ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 4, April 2023

Process Mining: An Emerging Tool for Data Science and Process Management

[1] Vibhuti, [2] Harshit Kumar, [3] Dr. Mahua Banerjee

[1] [2] [3] Student, Master of Business Administration, Birla Institute of Technology, Mesra (Lalpur Extension Centre), Jharkhand, India
Corresponding Author Email: [1] vibhuti897@gmail.com, [2] kumarharshit572@gmail.com, [3] mahuabanerjee@bitmesra.ac.in

***Abstract*— Process mining is a promising domain of data analytics. With an increasing orientation of businesses towards the Information Technology systems, process mining is creating strong roots for itself in process management. This paper has done a systematic analysis of various opportunities and perspectives offered by process mining. Also, the different stages, tools and techniques of process mining have been clarified. Apart from this, the paper discussed the positioning of process mining, data mining and business process management in data science. Though these concepts of data science are not new, how they are related to each other as well as how they differ from each other has been explained. To incorporate the application of process mining, the authors present the utilization of process mining in healthcare processes. This article further presents a clarified view for different techniques of process mining, viz. Process Discovery, Process Conformance and Process Enhancement; creation of event logs from healthcare records; utilization of event logs from Healthcare Information System (HIS) in creating process models; and application of the technique of Process Discovery, Process Conformance and Process Enhancement in understanding the efficiencies and the deficiencies in the real time healthcare processes. The paper finally discussed the present requirements in enhancing the efficiency of healthcare process mining.**

***Index Terms*—process-conformance, process-discovery, process-enhancement, process-mining.**

I. INTRODUCTION

Data mining is an essential and popular tool of data analytics. Process mining is gaining popularity because of its ability of process exploration and process intelligence. Identifying the slow and inefficient processes means identifying the bottlenecks within a process or a set of processes. The business processes seem simple, in general. But when analytics comes into play, one can easily understand the complexities of these processes. A process can have in itself multiple iterations of process units or sub-processes and within an individual iteration, there could be multiple conditions to be verified. Deviations and multiple interactions, nonetheless, creates equal complexities within a single business process. Process mining importance increases with the idea of determining the trends and patterns within a process or within different processes. Visualization of the explored processes creates a clear path to define those process units that can be automated or the process units that might need a substitution with a more efficient counterpart. Mining of business processes is, undoubtedly, a domain that requires more attention, especially with the current trend of an increase in the number of businesses managing their processes through Information and Technology systems. One of the most common IT systems that most of the businesses make use of these days is ERP. Process mining acts as a key in the improvement of the business processes. With the global impact of digital transformation, the approaches of process mining can be targeted towards digital analytics, AI-driven software for process mining and integration with other tools such as the cloud based tools. This article introduces process mining as a step towards data mining.

II. PROCESS MINING: BACKGROUND

Cook [9] is known to have published the first academic thesis on process mining. The technique of process discovery also appeared during this time. Later on the concepts of work flow mining, process modelling, business process management and other tools, techniques and algorithms were popularized by various group of researchers. van der Aalst [8] and their group is one of the major contributor in publishing process mining related research papers.

III. PROCESS MINING: MOTIVATION

Classical data mining techniques (such as classification, clustering, regression, association rule learning, and sequence/episode mining) do not focus on business process models and are often only used to analyse a particular activity in the overall process [1]. In the same aspect, the common traditional way of analysing the efficiency of business processes was done with the use of hand-made process models. The focal point of Process Mining, on the other hand, is on end-to-end processes. The term process mining can be understood as the technique of distilling a structured process description from a set of real executions [2]. Process mining does not start with an explicit process design, but aims at extracting process knowledge from "process execution logs" [3].

The possible creators of opportunities for process mining are:

A. Data Availability

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 4, April 2023

B. Discovery of New Processes
C. Techniques of Conformance checking

The use of Process Mining in the analysis and impoverishment of process flexibility again presents more opportunity for process mining as an efficient tool of data analytics. There are three dimensions that can be exploited for the classification of process flexibility. These are:

A. where does change occur (abstraction level of change)?
B. what is being changed (subject of change)?
C. how is it changed (properties of change)? [4].

IV. PROCESS MINING: PERSPECTIVES

There are mainly four different types of perspectives as a focal point of process mining. These are:

A. Control Flow Perspective

This is focused on creating an optimal path for process performance. The order of activities is used to find the optimal path.

B. Organizational Perspective

This perspective caters to create an optimal structure for various organizational units. It analyses the resources involved in a process, i.e., roles, departments, etc.

C. Time Perspective

The timing and the frequency of the events are focused in order to have optimal process timing and frequency.

D. Case Perspective

It is done by considering the properties that are present in different cases or types of processes. On further analysis the hidden dependencies and the relationship between the properties are revealed.

V. PROCESS MINING: STAGES

Process mining can be divided into four stages, namely

A. Activity

This stage involves creating a digital record of different activities, interactions and events.

B. Event Logs

This stage involves creating an event log from the digital records.

C. Visualization

This stage involves creating a visualized model of the process from the event logs.

D. Analytics

This stage involves analysis of the process, process comparisons and process conformance to diagnose the efficiencies and the deficiencies within the process.

VI. PROCESS MINING: TYPES

Wil van der Aalst, a Dutch computer scientist and professor, has contributed a lot in process mining through his academic research. According to him, Process mining is broadly classified into three types. These are discovery, conformance, and enhancement.

A. Process Discovery

This kind of process mining is mostly adopted. This technique primarily takes event log as an input and creates visualized process models as an output. Creation of these process models is possible using different types of process discovery algorithms, such as alpha algorithm, heuristic mining, genetic process mining and region based mining.

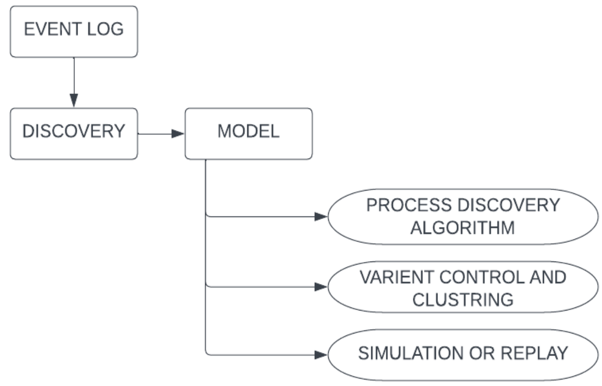

Figure 1. Process Discovery

B. Process Conformance

In this technique, the conformance of the reality is verified with an existing pattern to search for any deviations by doing a comparative study between the process model and the real process.

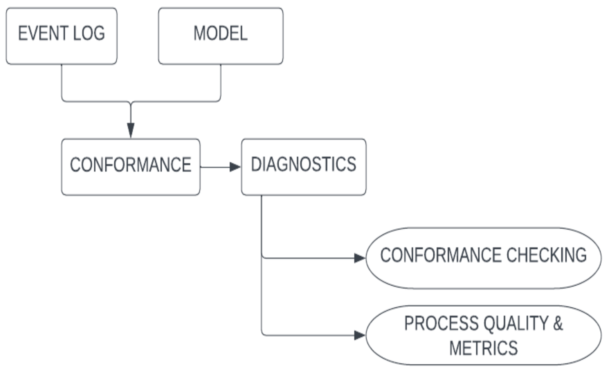

Figure 2. Process Conformance

C. Process Enhancement

The focal point of this technique is the enhancement of the performance of the process with additional attributes. These attributes could be data, such as, location, timing and cost.

ISSN (Online) 2394-2320
International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 10, Issue 4, April 2023

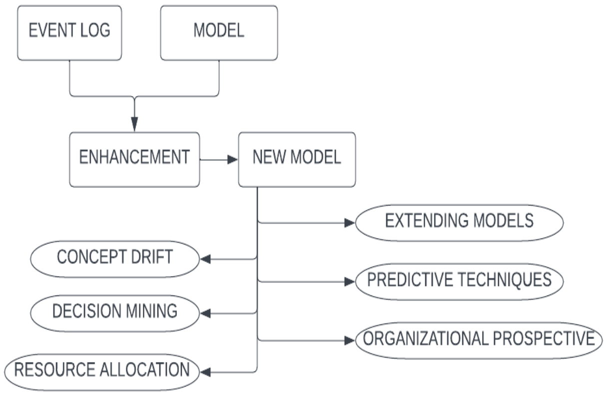

Figure 3. Process Enhancement

VII. DATA MINING, PROCESS MINING AND BUSINESS PROCESS MANAGEMENT

The active involvement of end-users, tool vendors, consultants, analysts, and researchers illustrates the growing significance of process mining as a bridge between data mining and business process modelling [1]. The relevance of business processes and the challenges associated with it are also presenting opportunities for an increase in the popularity of Process Mining in BPM (Business Process Management). Currently, many workflow vendors are positioning their systems as BPM (Business Process Management) systems [2].

VIII. PROCESS MINING IN HEALTHCARE

Process mining has gained an enormous importance in many sectors like IT Service management, Sales, Finance, Banking, Insurance, Healthcare etc. This is an emerging discipline providing novel techniques to infer valuable knowledge and information from event logs. The present BI tools not only focus on some querying and reporting together with visualization techniques but are also intelligent enough to convey some process mining capabilities. To explore the capabilities of process mining, the following section discusses the role of it in healthcare.

Just like any other information system, the Healthcare information system or HIS contains a lot of data in forms of tables. In addition, healthcare systems are confronted across the world with unprecedented challenges, including the permanent and rapid adaptation of clinical processes based on the emerging scientific evidence [6]. The healthcare data are related to information of the patients. These data can be utilized for the process analysis of healthcare to reduce cost. The records in the databases of these Healthcare information systems can be utilised to generate an event log describing the sequence of activities that were performed, during the execution of the events, by whom and for whom (e.g., considering a specific activity or a specific patient) [5]. Along with offering support to the data-driven management and improvement of healthcare processes, process mining offers potential opportunities to support the resilience of the healthcare system by facilitating a detailed analytical view to how processes are being executed within a particular context [6]. Process mining in healthcare can be exploited from different perspectives for identifying where and how to implement a quality improvement strategy to meet the healthcare needs.

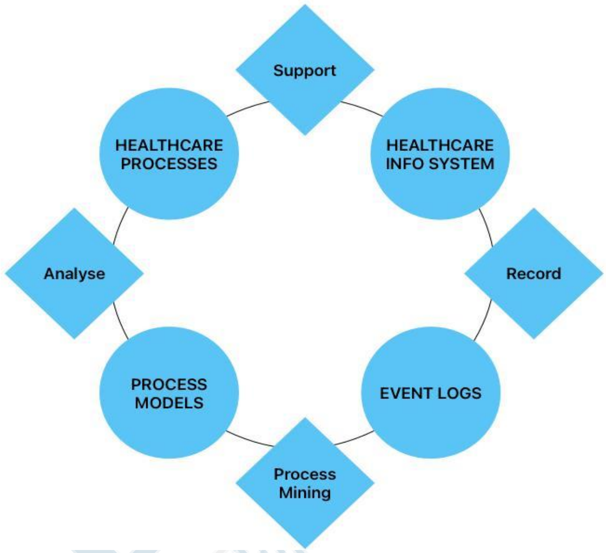

Figure 4(a). Process Mining in Healthcare

However, event log creation in healthcare environments faces considerable challenges due to heterogeneous data sources (e.g., mobile health data) and distributed healthcare providers [7].

As healthcare processes include diagnosis, treatment and disease prevention with the aim of enhancing patients' health hence process mining enables healthcare professionals to understand and analyze these processes in order to optimize the overall process efficiency.

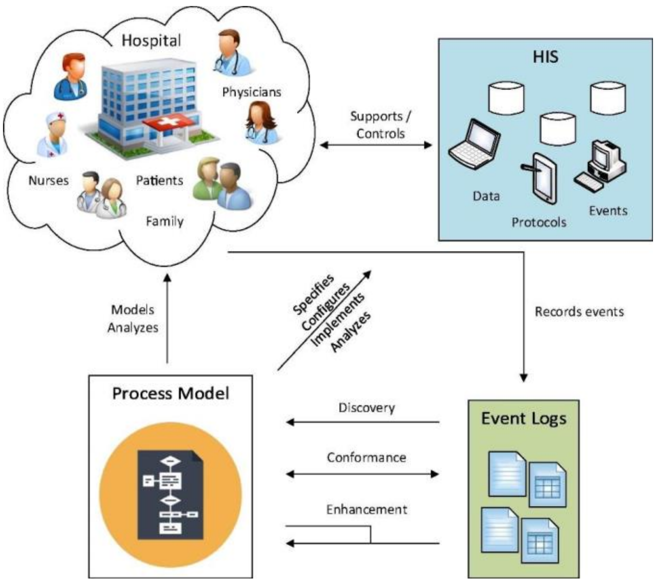

Figure 4(b). Process Mining in Healthcare [8]

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 4, April 2023

The above figure depicts the process mining activities in the healthcare domain. The HIS, the Event Logs along with the process mining techniques results in the process. These process mining models can cater to the problems of rising costs, population aging and increasing demand for healthcare. Moreover, the process mining technology provides insights about processes and patient behavior and help to drive efficiency by:

A. Identifying and understanding the real behavior of resources and the patients.
B. Analyzing process performance
C. Suggesting to redesign the process to reduce the waiting and service time, reduce the cost of services and to increase the process transparency
D. Detecting the bottlenecks in the process.

For example, Table I depicts the event log created for the diagnosis of 20 patients of thyroid blood test. These activities are encoded and shown in the third column of the Table I. Earlier, work has been done to show an example of creating event log for patients of gestational diabetes.

Table I: Event Log Example for Thyroid Blood Test

CASE	ACTIVITIES	ENCODING
PATIENT 1	<admission, blood test, check the result, sharing the result>	<a, b, c, d>
PATIENT 2	<admission, blood test, check the result, sharing the result>	<a, b, c, d>
PATIENT 3	<admission, blood test, check the result, request for additional test, additional test, check the result, sharing the result>	<a, b, c, e, f, c, d>
PATIENT 4	<admission, blood test, check the result, request for repetition, blood test, check the result, sharing the result>	<a, b, c, g, b, c, d>
PATIENT 5	<admission, blood test, check the result, request for repetition, blood test, check the result, request for additional test, additional test, check the result, sharing the result>	<a, b, c, g, b, c, e, f, c, d>
PATIENT 6	<admission, blood test, check the result, request for additional test, additional test, check the result, request for additional test, additional test, check the result, sharing the result>	<a, b, c, e, f, c, e, f, c, d>
PATIENT 7	<admission, blood test, check the result, request for additional test, additional test, check the result, request for additional test, additional test, check the result, request for additional test, additional test, check the result, sharing the result>	<a, b, c, e, f, c, e, f, c, e, f, c, d>
PATIENT 8	<admission, blood test, check the result, sharing the result>	<a, b, c, d>
PATIENT 9	<admission, blood test, check the result, request for additional test, additional test, check the result, request for additional test, additional test, check the result, sharing the result>	<a, b, c, e, f, c, e, f, c, d>
PATIENT 10	<admission, blood test, check the result, request for additional test, additional test, check the result, request for additional test, additional test, check the result, sharing the result>	<a, b, c, e, f, c, e, f, c, d>
PATIENT 11	<admission, blood test, check the result, request for repetition, blood test, check the result, sharing the result>	<a, b, c, g, b, c, d>
PATIENT 12	<admission, blood test, check the result, sharing the result>	<a, b, c, d>
PATIENT 13	<admission, blood test, check the result, request for additional test, additional test, check the result, request for additional test, additional test, check the result, request for additional test, additional test, check the result, sharing the result>	<a, b, c, e, f, c, e, f, c, e, f, c, d>
PATIENT 14	<admission, blood test, check the result, request for repetition, blood test, check the result, request for additional test, additional test, check the result, sharing the result>	<a, b, c, g, b, c, e, f, c, d>
PATIENT 15	<admission, blood test, check the result, sharing the result>	<a, b, c, d>
PATIENT 16	<admission, blood test, check the result, request for additional test, additional test, check the result, request for additional test, additional test, check the result, sharing the result>	<a, b, c, e, f, c, e, f, c, d>
PATIENT 17	<admission, blood test, check the result, request for additional test, additional test, check the result, request for additional test, additional test, check the result, sharing the result>	<a, b, c, e, f, c, e, f, c, d>
PATIENT 18	<admission, blood test, check the result, sharing the result>	<a, b, c, d>
PATIENT 19	<admission, blood test, check the result, request for additional test, additional test, check the result, request for additional test, additional test, check the result, sharing the result>	<a, b, c, e, f, c, e, f, c, d>
PATIENT 20	<admission, blood test, check the result, sharing the result>	<a, b, c, d>

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 10, Issue 4, April 2023

Here, a = admission, b = blood test, c = check the result, d = sharing the result, e = request for additional test, f = additional test, and g = request for repetition. The encoding of activities creates a simplified log of events. This event log can be utilized to create a process model.

Event log = [<a, b, c, d>^{7}, <a, b, c, e, f, c, d>^{1}, <a, b, c, g, b, c, e, f, c, d>^{2}, <a, b, c, e, f, c, e, f, c, d>^{6}, <a, b, c, e, f, c, e, f, c, e, f, c, d>^{2}, <a, b, c, g, b, c, d>^{2}]

The above equation depicts the aggregate event log of thyroid blood test of 20 patients. The aggregate event log is created from the data available in Table I.

The analysis of the aggregated event log further provides the scope of creating a process model using any of the process mining algorithms such as α-algorithm, Heuristic Mining, etc.

Process models can be further created for the different perspectives of process mining that are discussed earlier, i.e., control-flow perspective, organizational perspective, time perspective and case perspective.

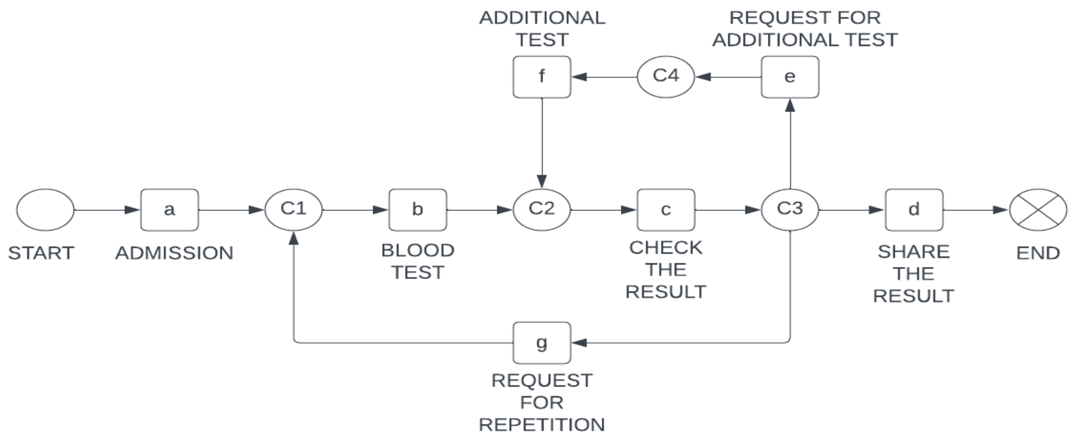

Figure 5. Control Flow Process Model using the α-Algorithm

Figure 5 shows the control flow model created from the event log of thyroid blood test using the α-algorithm. Here, activities are represented using squares and different stages are represented through circles. The circles are thus the connecting points for various transitions. This model can be utilized to create an optimal flow of different activities performed during a general thyroid blood test in a hospital or clinic by identifying the bottlenecks.

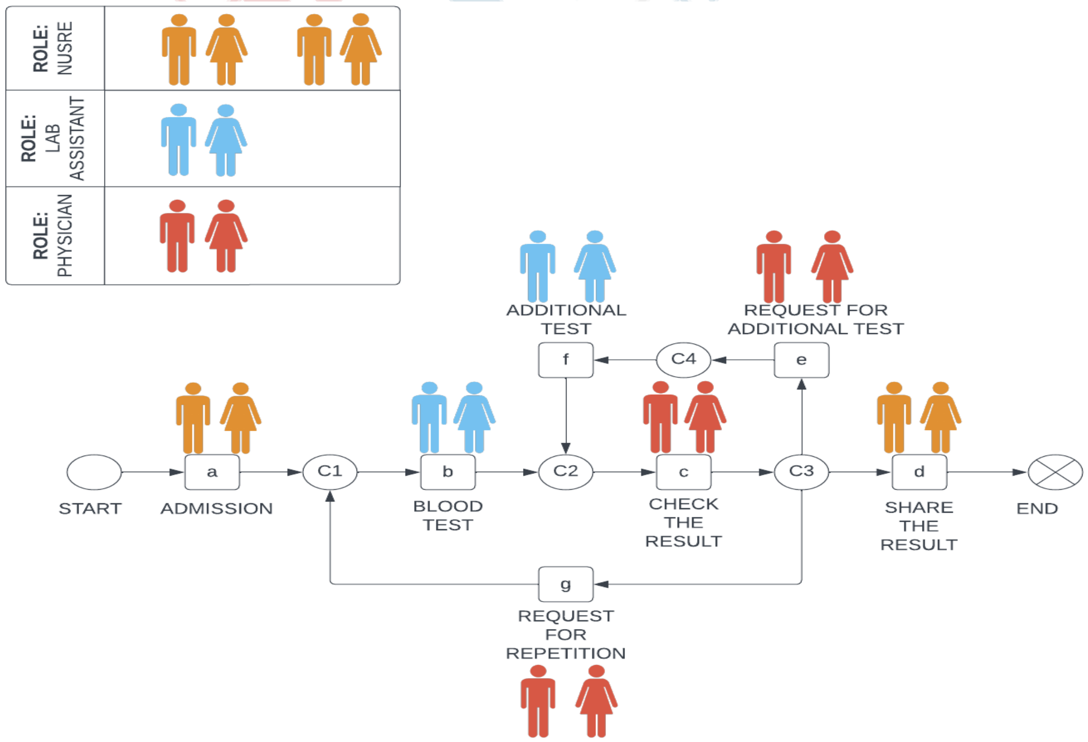

Figure 6. Process Model from Organizational perspective using the α-Algorithm

ISSN (Online) 2394-2320
International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 10, Issue 4, April 2023

Figure 6 represents the organizational model created using the α-algorithm and obtained from the event log. This model can be analyzed for better understanding of the roles played by different employees to perform the activities involved from the start to the very end of the complete process. This, thus, offers the opportunity to create an optimized organizational structure for an efficient process flow.

Figure 7 utilizes the time perspective of process mining to create a process model from the event log of the thyroid blood test in order to depict the time lag between the admission of a patient and his/her blood test and the time lag between the request for additional test and the actual performance of the additional test. Reducing or eliminating any unwanted time lag will result in more effective transitions.

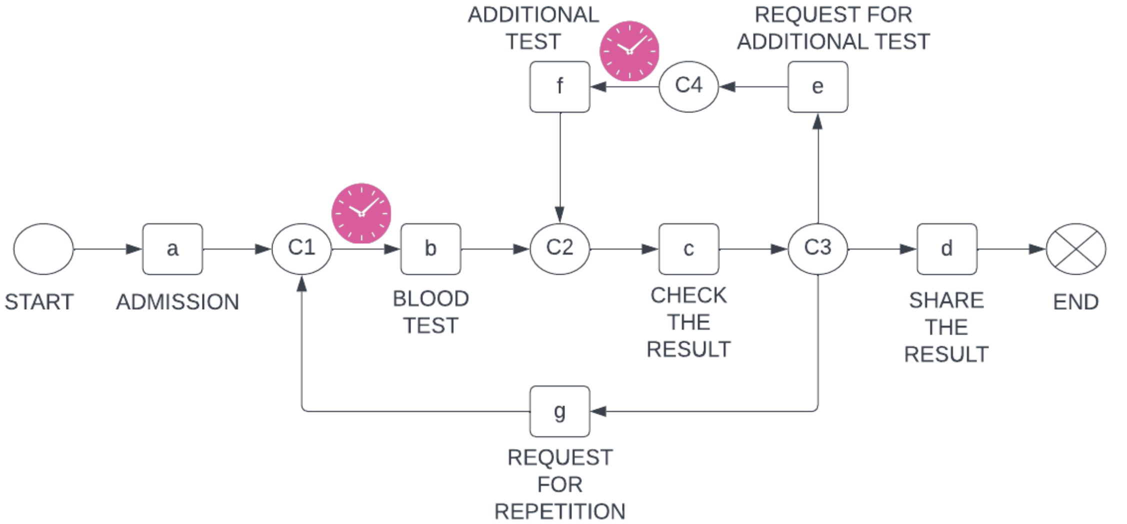

Figure 7. Process Model from Time perspective using the α-Algorithm

When the event logs present additional information that are specific for a particular case or two or more different cases, the process model can be created from the case perspective (Figure 8). For instance, the event log may have information related to the gender of the patients admitted for thyroid blood test. Analyzing the case related information, contribute towards a quick and accurate decision making activities associated with the process.

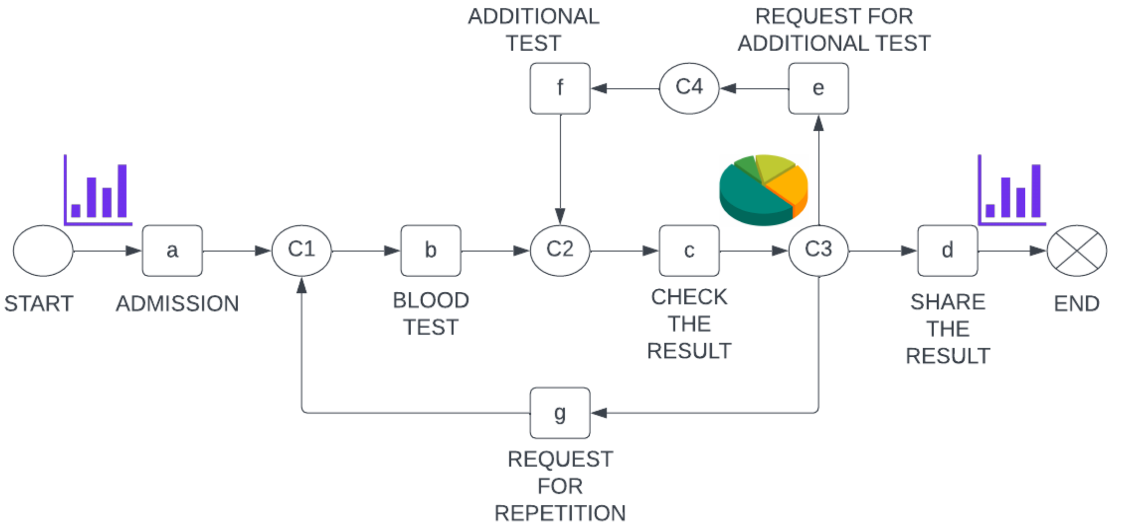

Figure 8. Process Model from Case Perspective using the α-Algorithm

IX. CONCLUSION

This paper has provided an overview of process mining by reflecting on the types, stages, perspectives and opportunities of process mining. The utilization of process mining is not only restricted in business organizations but in all sectors including healthcare. There are several challenges in mining processes in the healthcare domain which need utmost attention by the researchers. Healthcare processes are complex models due to different variations of treatments for the same disease, different conditions of patients and different ways and sequences in which the activities are performed by the resources like doctors, nurses, and other healthcare professionals. The healthcare process mining has been discussed to provide a holistic view of it. This will help in enhancing the healthcare process models to cater the dynamic processes of the system.

REFERENCES

[1] W. van der Aalst, "Process Mining: Overview and Opportunities," ACM Transactions on Management Information Systems, 3(2), 1–17, July 2012.

[2] W. van der Aalst, and A. Weijters, "Process mining: A research agenda, "Computers in Industry, 53(3), 231-244, June 2004.

[3] W. van der Aalst, C. Gunther, J. Recker, and M. Reichert, "Using Process Mining to Analyze and Improve Process Flexibility - Position Paper -," Business Process Modeling, Development, and Support, January 2006.

[4] G. Regev, P. Soffer, R. Schmidt, "Taxonomy of Flexibility in Business Processes," Business Process Modeling, Development, and Support, January 2006.

[5] W. van der Aalst, "Process mining: data science in action," Heidelberg: Springer, Vol. 2, 2016.

[6] N. Peiffer-Smadja, J. C. Lucet, G. Bendjelloul, L. Bouadma, S. Gerard, C. Choquet,... and F. X. Lescure, "Challenges and issues about organizing a hospital to respond to the COVID-19 outbreak: experience from a French reference centre," Clinical Microbiology and Infection, 26(6), 669-672, 2020.

[7] S. Remy, L. Pufahl, J. P. Sachs, E. Böttinger, and M. Weske, "Event log generation in a health system: a case study," In Business Process Management: 18th International Conference, BPM 2020, Seville, Spain, September 13–18, 2020, Proceedings 18 (pp. 505-522), Springer International Publishing, Chicago, 2020.

[8] W. Van Der Aalst, "Process Mining: Discovery, Conformance and Enhancement of Business Processes," Springer Science & Business Media, 2011.

[9] J. E. Cook, "Process discovery and validation through event-data analysis," Ph.D. thesis, Dept. Computer Science, University of Colorado, USA, Technical report CI-CS-817-96, 1996.