2020 International Conference on Intelligent Engineering and Management (ICIEM)

Anomaly Detection and Fault Prediction of Breakdown to Repair Process Using Mining Techniques

Rehet Bhogal
Department of Computer Science & Engineering
Amity School of Engineering & Technology
Amity University Uttar Pradesh
Noida, UP, India
rehetbhogal@gmail.com

Anchal Garg
Department of Computer Science & Engineering
Amity School of Engineering & Technology
Amity University Uttar Pradesh
Noida, UP, India
agarg@amity.edu

***Abstract*—A manufacturing company follows multiple business processes for the production of products. This paper uses process mining techniques on machinery breakdown and repair dataset obtained from a manufacturing company. The objective of applying process mining is to discover the runtime process model of the machinery repair process. It also focuses on assessing the as is process model from a control flow perspective, performance perspective and organizational perspective. The issues uncovered in the as is model are used to design a to be process model.**

Keywords—process mining; process discovery; process redesign; machinery breakdown

I. INTRODUCTION

A manufacturing organization aims at producing good quality products with high customer satisfaction at the same time maximizing profit. A manufacturing industry follows different processes for the transformation of raw materials into final products such as product assembly process, order to cash process, issue to resolution process, requisition process, quality assurance process, and corrective and preventive maintenance process. The corrective maintenance process, which deals with repair of machines, is particularly vital for a manufacturing company. The downtime of machines has a direct impact on the production and performance of a company. Inoperable machines do not contribute to production which renders the company unable to meet client demands. It not only incurs monetary losses but also damages the relationship with the stakeholders and company's reputation in the market. An organization suffers additional costs as time, money and resources are utilized in order to repair machines. Moreover, a machine might be a part of a sub-process of a bigger manufacturing process. In such a scenario, breakdown of a machine will stall the entire process. Hence, machinery breakdown and repair process requires critical attention.

During the breakdown and repair process, huge quantity of data is recorded by the organization's information system. This data may include information such as time of breakdown, type of breakdown, details of the machine, person involved in the maintenance of machine, type of fault and the steps followed to repair the machine, etc. This transactional data often contains hidden patterns and associations which cannot be interpreted manually. Automated analysis techniques such as statistical analysis and data mining help find correlations between various attributes and unveil interesting patterns that can be used for prediction and decision making by the organization and achieve competitive advantage.

This study uses process mining for analysis of machine breakdown and repair process. Process mining is used to obtain as-is process model, identify lacunas and bottlenecks in the system and subsequently redesign the business process.

The main contribution of this paper is to demonstrate how automated techniques such as process mining can be useful in the industrial domain for minimizing machinery breakdown so as to manage cost, time and efforts on the machine breakdown and repair process. The main motivation of the study is to improve the breakdown to repair process and discover patterns in machinery breakdowns so as to minimize the loss incurred to the organization and the unexpected downtime of machines and additionally improve the business process.

II. RELATED WORK

The research area which combines data mining techniques with business process management is referred to as process mining [1][2]. The purpose of process mining is extraction of information (process models) from event logs [3]. The omnipresence of hidden or actual event logs in information systems has made process mining a vivid research area. Before the emergence of this new research field, event logs were rarely used for the purpose of discovering underlying business processes. Process mining is basically a-posteriori analysis techniques for the exploitation of the information stored in event logs [4]. This is accomplished by using tools (eg, ProM framework [5], Disco [6]) and techniques for process model discovery, model conformance and model enhancement [7]. Process mining is a broad area in terms of its applications and techniques. Hence, process mining diagnoses the organization's processes by mining the event logs for knowledge extraction.

Process mining in relation to manufacturing industry was used by Aruna Devi. C et. al. [8] to analyse a manufacturing unit i.e., corrugated boxes manufacturing industry for process discovery and subsequent simplification of their processes.

978-1-7281-4097-1/20/$31.00 ©2020 IEEE

The event logs were analyzed with the alpha mining and the heuristic mining algorithm to construct a model. SookYoung Son et. al. [9] used process mining for the analysis of manufacturing processes by utilizing event logs from manufacturing execution systems (MES). Heuristic mining algorithm was used to derive the actual manufacturing process model. Conformance check and bottleneck analysis were performed which identified activities in the process with long lead time. Analysis from resource perspective gave insights into the relationship between activities and machine utilization. Mahendrawathi ER et al. explored customer fulfillment process to discover the process and assess the rate of a completed customer and time for processing different kinds of customer requests [10]. Process mining was also used in the healthcare industry by R. Mans et al. [11] for discovering the various clinical pathways undertaken by different hospitals. Different medical practices were discovered for treating similar patients and unexpected behavior was highlighted. C.W. Gunther et al. [12] have demonstrated the application of process mining for the monitoring of an X-ray system that is used by different devices. The event logs were used for discovering the actual application usage and the typical patterns followed using Fuzzy Mining. R.S. Mans et al. [13] used process mining techniques for discovering the typical paths followed by groups of patients in a Dutch hospital. The process was analyzed from control flow and performance perspectives and bottlenecks were detected. Arora and Garg [14] used process mining on logs recorded during bug resolution in Monorail Issue Tracking System. Different types and sources of inconsistencies were discovered in the resolution process. The extensive applications of process mining for process discovery and bottleneck identification gave motivation to use process mining for the study of breakdown repair process.

This study aims at discovering the breakdown and repair process of machines in a manufacturing company and also performing control flow and performance analysis to generate useful insights. Previous studies have used data mining techniques for predicting machinery faults on the basis of values recorded by sensors attached in machines. But our system does not have automatic sensor recorded data. The database of manufacturing company taken for the study, records the data manually by the employees. Hence, the type of data is not as extensive in previous studies.

III. Proposed Framework

The methodology consist of the following (1) Data collection and understanding (2) Data Preprocessing (3) Process Mining

1) Data Collection and Understanding: The data regarding breakdown and repair of machines was collected from a manufacturing company in Punjab, India. The company manufactures auto parts, twin wheels and cycle parts. Their products are sold not only in Punjab but also in several other states of India. A total of one hundred fifty two machines are employed for the manufacturing process such as slot milling machine, cup blanking press, cup grooving, etc. The company records information about breakdown and repair of machines in the database. The dataset obtained included a total of 6000

records of machinery breakdowns and repairs between the years 2012 and 2016. The dataset contains details regarding machine breakdown (Table 1) and the repair details of machines (Table 2).

A 'Slip_no' is generated when a breakdown complaint is recorded. The same 'Slip_no' is used to record the repair activities on the specific machine.

TABLE I. Machinery breakdown table description

S.No.	Attribute	Type	Description
1.	Slip No.	Numeric	The number of the slip which is generated when an employee registers a machinery breakdown complaint.
2.	Date	Date	The date when the complaint was recorded.
3.	Time	Character	The time when the complaint was recorded
4.	Emp No.	Numeric	The employee ID
5.	Machine No.	Numeric	The machine ID
6.	Process	Character	The process ID. It identifies the process for which the machine was being used.
7.	Fault	Character	This field records the reason why the breakdown occurred.
8.	DT Repair	Date	The date when the machine was repaired.
9.	TM Fitter	Character	The time when the fitter was assigned.
10.	DT Fitter	Numeric	The date when the fitter was assigned.
11.	Short At	Numeric	Previous process for which the machine was used.
12.	Short Due2	Numeric	The ID of the employee which previously used the machine.
13	Fitter	Numeric	The ID of the employee who was responsible for machinery repair.

TABLE II. Machinery repair table description

S.No	Attribute	Type	Description
1.	Emp No.	Numeric	ID of the employee who repaired the machine
2.	Slip No.	Numeric	This denotes the unique slip number of the slip that was generated on registration of complaint.
3	Dated	Date	The date when the machine was repaired.
4.	Time	Character	The time when the status of the machine was recorded after the machine was repaired.
5	Start Time	Character	The time when the repair of the machine began.
6.	End Time	Character	The time when the repair ended.
7.	Status	Character	The status of the machine after repair. It can hold two values – 'Finished' implying that machine was successfully repaired or 'Pending' implying that the repair was unsuccessful.

The ideal process for machinery breakdown and repair as described by the stakeholders is depicted in Fig. 1. The activities identified and their description is given in Table 3.

2) Data Preprocessing: The data was preprocessed so as to remove noise, missing values and inconsistencies in the recorded data. Once the data was preprocessed, the data was transformed into structured event logs for the purpose of process mining. The event logs serve as input for process mining techniques in order to extract non-trivial information about real-world processes.

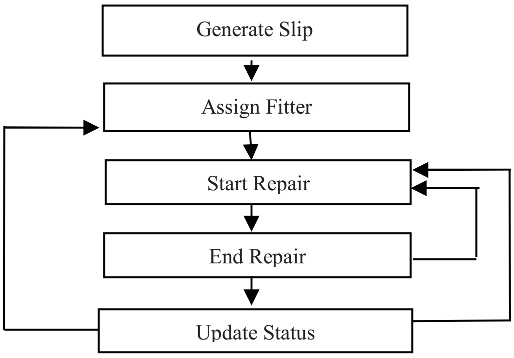


Fig. 1. The ideal process model for machinery repair

TABLE III. DESCRIPTION OF ACTIVITIES

S.No	Activity	Description
1.	Generate Slip	When an employee records a machinery breakdown complaint, a slip is generated.
2.	Assign Fitter	After a breakdown is recorded, a fitter is assigned to repair the machine.
3.	Start Repair	Beginning of machinery repair.
4.	End Repair	It denotes when the repair of machine finished.
5.	Update Status	The status of the machine after repair. It can take the values 'Pending' and 'Finished'.

a) Data Cleaning: There were several tuples that had no recorded values for the attributes *Fitter, Start_time, End_time*. The missing *Start_time* and *End_time* were calculated using the attribute *Time_spent* and manually filled. The attribute *Slip_no* is supposed to store unique values. However, it was found that several slip numbers had been repeated twice and such tuples were allotted new slip numbers. The dataset consisted of two different data types for storing time. There was a 24- hour time format with column entries as *HH:MM* and a 12-hour time format stored as decimal numbers up to two places. Such column entries were in the form *HH.MM*. The columns storing time were converted to 24-hour time format *HH:MM.*

b) Data Integration: While performing integration on the dataset, entity identification problem and tuple duplication had to be dealt with. Entity identification problem addresses the issue of matching equivalent real-world entities contained in several data sources.

c) Data Reduction: Data was reduced using dimensionality reduction. A subset of attributes relevant to the study was selected.

CASE_ID	TIMESTAMP	ACTIVITY	RESOURCE
3365	22/01/2013 10:09:00	ASSIGN FITTER	2435
3365	28/01/2013 10:15:00	END REPAIR	2101
3365	22/01/2013 09:32:00	GENERATE SLIP	2603
3365	28/01/2013 08:30:00	START REPAIR	2101
3365	28/01/2013 15:52:00	UPDATE STATUS	2101
3366	22/01/2013 09:46:00	ASSIGN FITTER	2435
3366	22/01/2013 17:00:00	END REPAIR	2101

d) Data Transformation: During data transformation, the data set was transformed into event log. Data from both tables was transformed into event logs which are presented in Fig 2.

Fig. 2. Data after transformation into event logs

3) Process Mining: Process mining primarily focuses on discovering, monitoring and improvement of real processes through the extraction of knowledge from event logs.

a) Process Discovery: After preprocessing, 5036 cases were extracted. The event logs were imported into Disco (process mining tool) and as-is process model showing directed map of activities was discovered. This model depicted the actual process model being followed by using the actual activities recorded in event logs.

b) Process Analysis: The discovered process model was analyzed from control flow perspective and organizational perspective [15]. The former analyzed the variants of the model and its performance. The variants depicted the deviations from the ideal flow. The performance evaluation of the model revealed the duration between activity transitions. Organizational perspective analyzed the resources involved in the process. It gives information regarding performance of the resources, the frequency of resource participation.

IV. EXPERIMENTAL RESULTS & DISCUSSION

A. Process Discovery

The real time process was discovered using Disco tool and is shown in Fig. 3. A total of 5036 cases, 5 activities and 25,999 events had been recorded *between 21-04-2012 and 10-09-2016* with mean case duration of 9 days. The thicker the transition lines are, the more are the number of cases that followed that particular transition in the direction of the arrow. It is evident from the diagram that a considerable number of cases followed the ideal process. However, there are numerous other paths which were undertaken revealing that the process actually followed by the workers does not comply with the one to be followed.

B. Variant Analysis

For a total of 5036 cases, 112 variants were discovered and only 3 variants out of them belonged to the ideal flow. The variants were studied and were categorized in vital groups in Table 4.

1) Early Status Update: The exceptional flow depicted in Fig. 4, wherein the status of the machine was updated before the repair work concerns resource behaviour. This flow makes up 22.7% of the entire process. It is more prevalent in cases where there is a gap between allotment of fitters and the day of

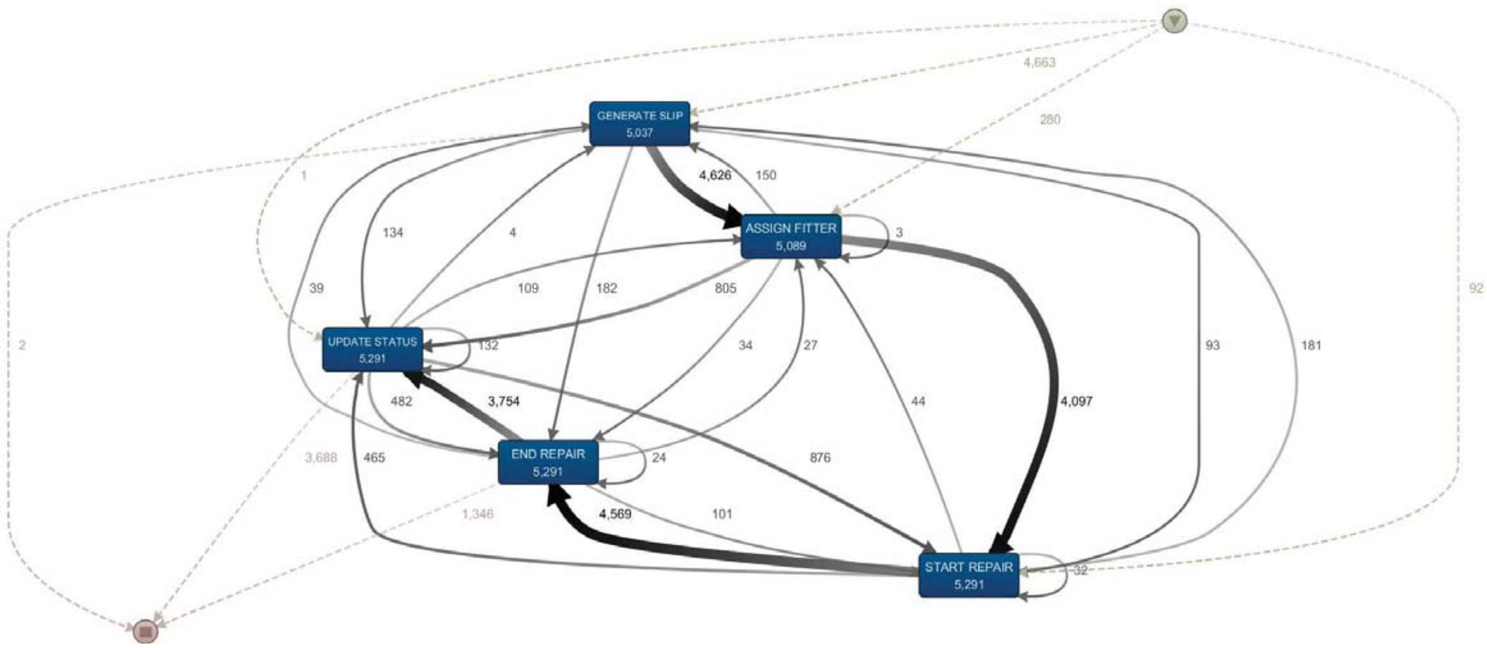


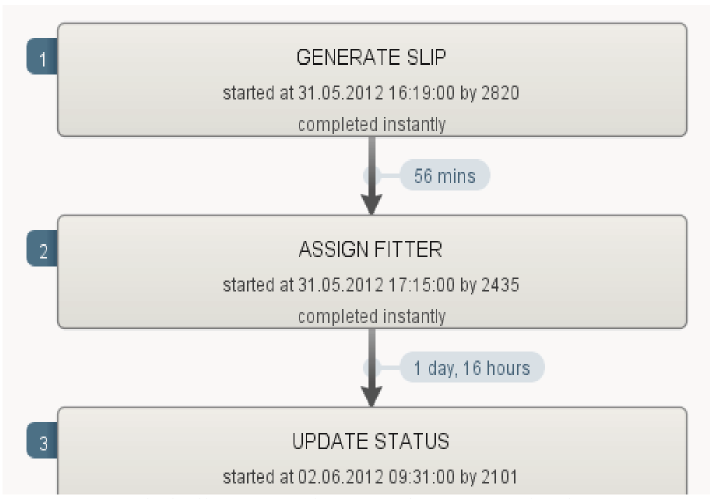
Fig. 3. The discovered process model with absolute frequencies

machinery repair. There should be a check in the system that status of machines can be recorded only after the machine has been repaired and if an attempt is made to do so, system should notify the administrator.

S.No	Type	Cases	Percentage
1.	Ideal Process	3296	66%
2.	Early Status Update	1144	22.7%
3.	Delayed Slip Generation	299	5.9%
4.	Rework Required	63	1.2%
5.	Fitter Reassigned	27	0.5%

TABLE IV. MAJOR CATEGORIES OF VARIANTS AND THEIR PERCENTAGE

2) Delayed Slip Generation: The process should begin with generation of slips. However, 299



cases recorded slip generation at a later stage in the process; even as late as when the status of the machine was reported. This exceptional flow is shown in Fig. 5. With the presence of such anomalies, one cannot analyze resource efficiency accurately.

3) Rework Required: 1.2% of the cases required rework up to 4 times till the repair was successful. Some of the repairs were scheduled at a significant gap. Although this is not a very big number, but it affects the cost and productivity. For the purpose of rework, 27 cases required the fitter to be reassigned as the previous workers did not successfully finish their job. Ideally, machine should get

S.No.	Case Duration	Cases
1.	Upto 13 days	4627
2.	Upto 26 days	166
3.	Upto 39 days	72
4.	Upto 52 days	71
5.	Upto 65 days	25
6.	Upto 78 days	10
7.	Upto 91 days	6
8.	More than 100 days	60

repaired within the same day.

Fig. 4. Early status update variant extracted from Disco tool

Fig. 5. Delayed slip generation variant

C. *Process Performance*

The performance of a process can assessed by analysing the duration of process instances, individual activities and time between transitioning from one activity to another. The duration of cases is discussed in Table 5. Majority of repairs finished within 13 days. This is not a good enough number from performance perspective. Delays in repairs hamper the productivity of the company. Case duration should be brought down to atleast a week. In addition, a priority should be assigned with each machine repair. This will facilitate analysis of machines based on priorities.

TABLE V. DURATION OF CASES

Machine breakdown and repair process was redesigned to eradicate the abnormal flows. Further, problem areas with the ideal flow were identified. Table 6 presents the performance of the process for the ideal activity transitions with the mean duration between these transitions. In certain rework scenarios, a new fitter was assigned after the previous fitter was unable to successfully repair the machine. This is depicted with the activity transition *Update Status -> Assign Fitter*. It is observed this process took a mean duration of 9.4 days which is a concern. The mean duration between the assignment of a fitter and when the fitter started to repair the machine is 4.3 days. This suggests that the process of machinery repair was delayed by the fitters. In case of rework by the same fitter, a mean time of 37.9 hours was taken to start with the machinery repair work again. Once the repair work began, a mean duration of 2.5 hours was taken to complete the work which is acceptable. Hence, most of the time in machinery breakdown and repair process is consumed in reassigning fitters, rework on machines by fitters and starting repair work by fitters.

TABLE VI. PERFORMANCE OF THE PROCESS

S.No.	Event Transition	Mean Duration
1.	Generate Slip -> Assign Fitter	21.1 hours
2.	Assign Fitter -> Start Repair	4.3 days
3.	Start Repair -> End Repair	2.4 hours
4.	End Repair -> Update Status	2.5 hours
5.	Update Status -> Start Repair	37.9 hours
6.	Update Status -> Assign Fitter	9.4 days

D. *Resource Analysis*

Resources play a key role in a business process model of an organization as they are responsible for carrying out various operations. Resources can either be human workers or machines. In our case study, workers are the resources responsible for carrying out various tasks. The completion and duration of tasks depends copiously on the resources. The mean duration of a worker with respect to every machine he repaired was analyzed and gathered. The total number of combinations was found to be 769. The overall resource efficiency is presented in Table 7. It shows the mean duration of resources between the activities *Start Repair -> End Repair*, i.e., the time taken to repair machines.

TABLE VII. OVERALL RESOURCE EFFICIENCY

S.No.	Resource	Cases	Events	Mean(hrs)
1.	3300	355	746	5.7
2.	3264	38	76	2.6

3.	3214	171	352	3.1
4.	3210	7	14	1.6
5.	3171	19	40	2.9
6.	3153	864	1780	18
7.	3122	70	140	3.3
8.	3047	31	62	2.7
9.	2862	236	482	3.6
10.	2712	48	96	3.1
11.	2655	233	470	2.8
12	2616	7	14	2.5
13.	2437	67	136	2.1
14.	2370	131	264	3.3
15.	2361	87	176	3.5
16.	2284	50	100	2.4
17.	2260	1128	2448	6.2
18.	2101	1268	2602	5.8
19.	1507	150	308	5.1
20.	1304	134	280	3.2
21.	509	4	8	3

It is observed that resource 2101 worked on majority of the cases, i.e., 1268 cases, 746 events/activities and took a mean duration of 5.8 hours to repair machines. The longest mean duration of 18 hours for repairing machines belongs to the resource 3153. The mean duration for most of the resources lies between 2 – 3.5 hours.

V. Conclusions

This study depicts the way in which mining techniques can be used for identifying abnormal conditions. It demonstrated the use of process mining for detection of anomalies in the machine breakdown and repair process. Data obtained from a manufacturing company was transformed into event logs to discover the as-is process model of machinery breakdown and repair process in Disco. The model was analyzed from control flow and organizational perspectives. The manufacturing company can manage cost, time and efforts by analyzing the event logs to record abnormal behaviour.

References

[1] W. M. P. van der Aalst, H. A. Reijers and H. M. W. Verbeek, “Business process mining: An industrial application,” *Information Systems*, vol. 32, pp. 713-732, Jul. 2007.

[2] M. Pospisil, T. Hruska, V. Bartik and V. Mates, “Process Mining in a Manufacturing Company for Predictions and Planning”, *International Journal on Advances in Software*, vol. 6, 2013.

[3] W. M. P. van der Aalst and A. J. M. M. Weijters, “Process mining: A research agenda”, *Computers in Industry*, vol. 53, pp. 231-244, 2004.

[4] C. W. Gunther , S. Rinderle, M. Reichert and W. M. P. van der Aalst, “Using process mining to learn from process changes in evolutionary systems,” *International Journal of Business Process Integration and Management,* vol. 3, pp. 61-78, Jul. 2008.

[5] B. F. van Dongen, A. K. A. de Medeiros, H. M. W Verbeek, A. J. M. M. Weijters and W. M. P. van der Aalst, “The ProM framework: A new era in process mining tool support,” *Lecture Notes in Computer Science*, vol. 3536, pp. 444-456, Jun. 2005.

[6] *Disco.* (2017). Fluxicon. Available: http://fluxicon.com/

[7] Mekhala, “A review on process mining,” *International Journal of Engineering and Techniques*, vol. 1, pp. 11-17, Jul. 2015.

[8] A. D. C and Sudhamani, “Application of business process mining using control flow perspective in manufacturing unit,” *International Journal of Emerging Trends & Technology in Computer Science*, vol. 2, 2013.

[9] S. Son et. al., " Process mining for manufacturing process analysis: A case study," in *Asia Pacific conference on Business Process Management 2014*, Brisbane, Australia.

[10] M. ER, H. M. Astuti and A. Nastiti, “Analysis of customer fulfilment with process mining: A case study in a telecommunication company,” *Procedia Computer Science,* vol. 72, pp. 588 – 596, Dec. 2015.

[11] R. Mans, H. Schonenberg and G. Leonardi, “Process mining techniques: An application to stroke care,” *Studies in Health Technology and Informatics*, vol. 136, pp. 573-578, Feb. 2008.

[12] C. W. Gunther, A. Rozinat, K. V. Uden and W. M. P. van der Aalst, “Monitoring deployed application usage with process mining,” BPMCenter. org, Eindhoven, NB, Netherlands, BPM reports, 2008, vol. 811.

[13] R. S. Mans, M. H. Schonenberg, M. Song, W. M. P. van der Aalst and P. J. M. Bakker, “Application of process mining in healthcare – A case study in a Dutch hospital,” *Communications in Computer and Information Science*, vol. 25, pp. 425–438, Jan. 2008.

[14] A. Garg and R. Arora, “Analysis of software repositories using process mining,” *Smart Computing and Informatics*, vol. 2, pp. 637-643, 2018.

[15] W. M. P. van der Aalst, et al., "Process mining manifesto," in *Business process management workshops*, Springer, Berlin, Heidelberg, 2012, pp. 169-194.