Process mining case study approach: Extraction of unconventional event logs to improve performance in Hospital Information Systems (HIS) Corresponding Author

6 authors, including:

Gopi Battineni
University of Camerino
19 PUBLICATIONS   **3** CITATIONS

SEE PROFILE

Nalini Chintalapudi
University of Camerino
10 PUBLICATIONS   **2** CITATIONS

SEE PROFILE

Vania Karami
University of Camerino
11 PUBLICATIONS   **4** CITATIONS

SEE PROFILE

Giulio Nittari
University of Camerino
23 PUBLICATIONS   **38** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Breast cancer tumor model View project

PROCESS DISCOVERY AND CONFORMANCE CHECK BY PROCESS MINING ON MEDICAL TRAINING DATSET FOR SURGICAL PROCESS View project

Process mining case study approach: Extraction of unconventional event   logs to improve performance in Hospital Information Systems (HIS)

GOPI BATTINENI[1*], NALINI CHINTALAPUDI[2], VANIA KARAMI[3], GIULIO NITTARI[4], FRANCESCO AMENTA[5], S K TAYABATI[6]

[1, 3, 4] Doctoral candidates at Telemedicine and Tele pharmacy, School of Pharmaceutical Sciences and Health Products, University of Camerino, Italy

[2] A master student in computer science from JNT University, India

[5] Professor and president of tele pharmacy Pharmaceutical Sciences and Health Products, University of Camerino, Italy.

[6] Professor of Cellular Biology and Human Anatomy, University of Camerino, Italy

Corresponding Author
Dr. Gopi Battineni
Research Student on Health Informatics
University of Camerino
Telemedicine and Tele-pharmacy Center Via Madonna Delle Carceri 9, 62032 Camerino (Italy)
Email: gopi.battineni@unicam.it

ABSTRACT:

Background: Due to recent advancements in the computing technologies, process-mining algorithms can consequently apply complex numerical calculations in big data analysis with quick repetitions. It involved primarily in real-time systems such as banking, transport, and hospitals. However, process mining tending to play a vital role in the healthcare domain with promising results.

Objective: The goal is to extract the unconventional event log file from the process to improve system performance and to propose solutions to enhance system performance

Methods: The present research incorporated the performance filter technique to sort out misbehaved event logs from the process. We consider a sample of 100,000 patient cases over three-years from the local Hospital information systems (HIS).

Results: Results were emphasized with precise coordination, performance assessment by specific mining tool has done. Percentage of time taken to process unconventional event log files is validated as 63.67%, which means system performance getting low due to these events

Conclusion: In particular, to gain competitive advantage hospitals are trying to streamline their processes through an accurate view of the care flows. Several studies were done to integrate these techniques with hospital software systems, and we are come up with the contemporary idea of event log filtering

Keywords: Process mining, Process discovery, Event logs, HIS, and Process enhancement.

1. Introduction

Process Mining is a contemporary knowledge of data mining that can highlight among the original articles and focuses on extracting knowledge from available information stored in the databases of IT systems [1]. In recent times, many studies were conducted to analyze hospital processes, including Business Process Redesign, Evidence-Based Medicine [2], and among others.

Besides, process mining can be done in three methods through process discovery, conformance checking, and enhancement. Each technique has its strategies and protocols. Process discovery allows for model extraction from an event log file; conformance check will monitor data deviations between the generated model and the actual model, and finally, enhancement uses originally recorded event data to extend or improve an existing process model [3]. Currently, these techniques were in high demand for healthcare [4] systems due to significance for methods to figure out the clinical pathways and the healthcare service redesign [4], [5].

Despite that, healthcare systems designed by complex models with the subject to significant variation over time[6] and these variations will be caused by multiple factors such as performing resources at different patient conditions or activities. At present, many studies were attempted with process mining to discover process flow. Chapela.et.al [7] proposed an approach for searches process model to extract patterns, loops that usually executed in event logs. Authors present Wo-mine algorithm to recover common behavioral patterns from the model. Likewise, other studies proposed a methodology for business process analysis in hospital systems [8] and implement a tool, which integrates the primary stages of process analysis. Also, another study conducted on identifying causes to delay in healthcare refund process [9] of the medical institution in Brazil. They extract necessary comprehensions relate to processing execution and confirm the process efficiency and ability of process mining combination and accompanying rules in mining techniques.

Similarly, the mentioned study was proposed a novel methodology for extracting unconventional event log files from process discovery in hospital information systems (HIS) and visualize their time delays. We address performance filter techniques, which is one of the critical planning to define unusual activities in the process — the rest of the paper structured as follows. Section 2 will describe the research background and description of the key terms involved in process mining. Part 3 will present a methodology developed in the case study, and section 4 will depict the results about how mentioned mining tool helping to achieve the proposed method. Finally, the paper will end with a small conclusion note through chapter 5.

2. Research background

This section presents the basic terminology used in process mining related to the scope of this study.

2.1. Process Mining

About fifteen years ago, Cook and Wolf were attempted first process mining practical applications. Eventually, several works are conducted in a field that focused on the development of new algorithms for the automatic discovery of business processes [10]. After that, process mining was developed as a young research discipline to achieve the appropriate extraction that uses an event log data containing all the executed activities for support, control and further process.

In general, when software systems interact with the real world, event data will be created. The event data happen at a particular time stamp, and collection of events known as the "event log." Also, these event logs might contain event identification information, Name, timestamp, and others. Process Mining has done with this event logs, and these techniques help to handle to operate real-time cases [11]. Therefore, process mining can be considered as a connection between event log data and process model that used to configure software systems [12].
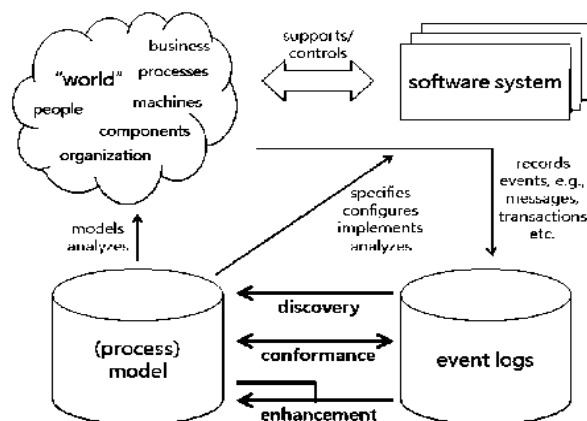

Figure 1. The graphical view of how process mining will work in a real-time world [13].

2.2. Scope of process mining in HIS

HIS had designed for storage of patient data in event log format by human acknowledgment. These event logs processed for enhancing different goals of medical centers such as financial, staff monitoring, and mainly on patient services [14]. By clustering algorithm, large event logs are splitting into subparts (i.e., events with the same attributes) which will emphasize the service quality in medical centers. Also, to address communication issues these methods enabling the frequent interactions between HIS activities. Therefore, process mining regularly comes up with the right approach and these methods might adjust or even dispense with an ultimate objective of change.

Many studies generate different methods to healthcare processes for discovering process models from event logs. Proposed techniques or algorithms will provide an expert system in medical operations and the ability to respond to the questions about performance issues, thereby generating improvement opportunities [16]. The ability to discover process model and performance analysis provides an excellent opportunity for data stored in event logs of HIS. Application of these techniques in HIS not only ensures performance improvement but also benefited in service quality improvement and a positive impact on the management of medical centers [17]

However, to achieve an advantage of information stored in HIS event data, there is a need to utilize techniques for process model discovery, performance analysis. Research outcomes would be beneficial for all hospital staff to make accurate decisions and explains the workflow of the system including process models with visual analytics. Eventually, process mining can visualize the real process flow steps in the system. Executed process analysis of health systems needs knowledge extraction from a stored data set since differences in patient conditions, and healthcare dataset consists of significant variation over time [15]. Process mining utilized for optimization of unusual behavior of event logs based on the information of HIS Database.

3. Proposed Methodology

The case study has extracted from 4TU.research data center [18] with 100,000 cases over three years. Event logs were created by business operations of the local hospital and each log records hospital activities for execution of billing on medical services. The concern is system performance was getting low since some logs were having much delay when they are moving from one activity to another. The study objective is to contemplate the reasons behind the erratic behavior of such log files and offer solutions. Figure2 will legitimize the systematic methodology involved in the study.
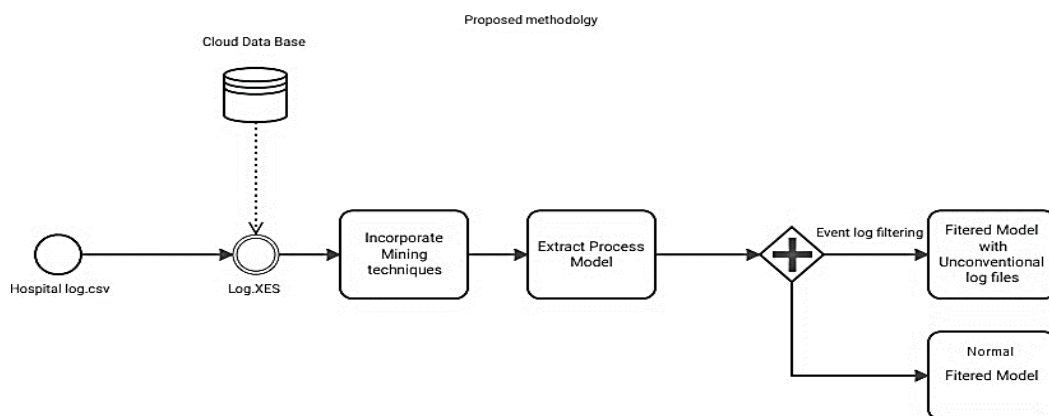

Figure 2 proposed mining methodology incorporation of model discovery and long-run event extraction

3.1.   Event log observation from data

Data is the starting point of any process. This data is measured and calculated by human knowledge, and these conditions were available on XES standard [3]. In general, data files accessible in CSV format; however, most of the process mining tools do not accept CSV type files. Therefore, it is necessary to convert CSV to XES for extracting event log behavior

3.2.   Incorporate mining techniques

Data preprocessing is one of the essential and mandatory elements in data mining because of most of the data is available in the real world is a rare and missing value type. So, such missing values are validated with the average of previous entries.

3.3.   Model Extraction

As discussed earlier, model discovery is one of the novel approaches in process mining, and it will enable users to visualize how the process was done and predict event log behavior[3], [4]. It is prerequisite to perform observation of event log flow that will help to filter out misbehaved events, which tended to affect the system performance. By incorporating the filtering methods with the process model, the synthetic view of activities followed by the hospital departments has determined. Process discovery is happened by importing log file into a unique mining tool (i.e., disco) and apply mining approach

3.4.   Event log filtering

Filtering is a prominent approach that will able to find out high-quality models without any deviation. This will also enable us to concentrate on a specific part of the process or particular case types. Therefore, we focus on the long-running cases (≥ 380 Days) from the process and generate an outcome model with long running ones.

4. Preliminary analysis and discussion

4.1.   Process discovery

After importing event log file, the system will automatically validate the process discovery step and generate the process model (figure 3) included with all the events. Total 100,000 cases are performing with 18 activities, and the mean duration of each activity is validated with 18.2 weeks (Table 1). All the cases took about three years to complete the entire process. Cases begin at Triangular symbol (i.e., starting point) in the process and reach the end (i.e., square symbol). The model containing all activities with case frequencies and mean duration has observed. The thick red paths are representing a constant delay in case flow between operations.   The extracted model is involved in 18 activities. However, activities with high-frequency events were only CHANGE DIAGNOSIS, FINANCE, RELEASE, CODE OK, AND BILLED (Refer table 2), which engaged with a great exchange of event data. Observations were interpreting that few cases do not reach the end. Such behavior may cause due to the mismatching of event data and stored data from HIS.

Table1: Statistics of the eventual process model

Parameter	Static Value
Events	451,359
Cases	100,000
Activities	18
Mean Case Duration	18.2 weeks
Starting timestamp	13.12.2012 10:13:18
Ending time stamp	19.01.2016 08:58:56

As mentioned before, the total period to finish the process is observed about three years such that high-frequency case activities are performing long delays compared to others. We set a threshold value 380 days (1 year and two weeks) that the cases took more than threshold time called as low-performance activities. Overall, due to the time elapse caused by the event overflow, the happening of highest delays of 36 weeks for case exchanging from DELETE to JOIN-PAT and 25.4 weeks for NEW to JOIN-PAT activities was mostly affecting the process performance.

4.2. Incorporation of Performance Filter

To investigate the misbehavior of some event logs, we incorporated the 'Performance filter' into the outcome model. Frequency of events and percentage was evaluated from table 3. Outcomes mentioned that only two percent of event logs (Table 4) are performing a radical manner (i.e., long-running) in the process that was extending the process execution time. Outcome model with long-standing cases is illustrated in figure 4. From the outcome model, we had noticed the high amount of delay is happening between DELETE to JOIN-PAT and NEW to JOIN-PAT as approximately equal to 19 months and 18 months respectively. Besides, a significant backlog of 47 weeks for case exchange between BILLED to MANUAL activities. Eventually, event data flow at these activities accomplished with unpredictable behavior corresponds to others.

Consequently, the staff at these activities working with event data that will not help in the growth of the system performance. By avoiding such delays that will automatically improve the HIS performance value. In conclusion, proposed methods can assist in raising the process performance and provide immediate suggestions to hospital authority by reporting about data deviation at particular activities.

Table 2: activities with the absolute frequency values

Activity	Event Frequency	Relative frequency
BILLED	67,448	14.94%
CHANGE DIAGN	45,451	10.07%
CHANGE END	38	0.01%
CODE ERROR	75	0.02%
CODE NOK	3,620	0.80%
CODE OK	68,006	15.07%
DELETE	8,225	1.82%
EMPTY	449	0.10%
FIN	74,738	16.56%
JOIN-PAT	358	0.08%
MANUAL	372	0.08%
NEW	1,01,289	22.44%
REJECT	2,016	0.45%
RELEASE	70,926	15.71%
REOPEN	4,669	1.03%
SET STATUS	705	0.16%
STORNO	2,973	0.66%
ZDBC_BEHAN	1	0%

Table 3: Frequency of events after implementation of performance filtering method

Activity	Event Frequency	Relative frequency
BILLED	3,502	16.31%
CHANGE DIAGN	1,141	5.31%
CHANGE END	3	0.01%
CODE ERROR	22	0.10%
CODE NOK	385	1.79%
CODE OK	3,778	17.59%
DELETE	144	0.67%
EMPTY	35	0.16%
FIN	3,374	15.71%
JOIN-PAT	70	0.33%
MANUAL	58	0.27%
NEW	2,893	13.47%

REJECT	836	3.89%
RELEASE	3,478	16.20%
REOPEN	586	2.73%
SET STATUS	201	0.94%
STORNO	968	4.51%

Table 4: Statistics of process model after performance filtering

Parameter	*Static Value*
Events	21,474
Cases	2,875
Activities	17
Mean Case Duration	15.9 months
Starting timestamp	13.12.2012 23:35:24
Ending time stamp	19.01.2016 10:58:56

4.3.   Performance Assessment

Performance of the system is the percentage of time taken to finish the formula calculates the entire process and rate of time taken to process conventional event logs

$$\frac{}{\text{number of days to complete process}} \times 100, \text{ i.e. } \frac{380}{1046} * 100 \text{ nearly about 36.3\%.}$$

A number of days to finish to complete the normal events a Total process

Similarly,   Percentage   of   time   taken   to   process   radical   event   log   files $\frac{\text{Rest of the days to finish the process after normal event extraction}}{\text{Total number of days to complete the process}} \times 100, \text{ i.e. } \frac{666}{1046} \times 100$ nearly   63.67%.

Therefore, it is evident that particular process is taking a long time to resolve the issues of long-running cases and by extracting unconventional event log files, it possible to improve the system performance of about 64%.

5. Conclusion

Mentioned work discusses the case study of local hospital billing information. To finish some activities on time, we performed some conventional process modeling techniques. The results also stimulated with the system performance because some event logs in the HIS are having unpredictable behavior. Evaluation of a system especially with big data needs an organized plan that can be assessed on an annual basis. Until now, health care organization decisions are based on paper reports that will not be considered as reliable sources to decision plan. Lack of implementing process mining might tend these systems to worse services with low quality, and

decreasing the efficiency of the services instead of improvement. Therefore, authorities should include process mining results in their management plan for decision- making. The proposed methodology can use at rearranging the hospital units or even at designing new hospitals at early phases. Furthermore, this is a comprehensive model for all health-related area, business information exchange approaches. Since the process, mining has covered a variety of applications; this can implement any service-oriented enterprise.

Conflict of Interest: Authors do not have any conflict of interest

Author Note: Present paper is cited and checked by all the authors and prepare the presented manuscript with error-free and grammatically correct
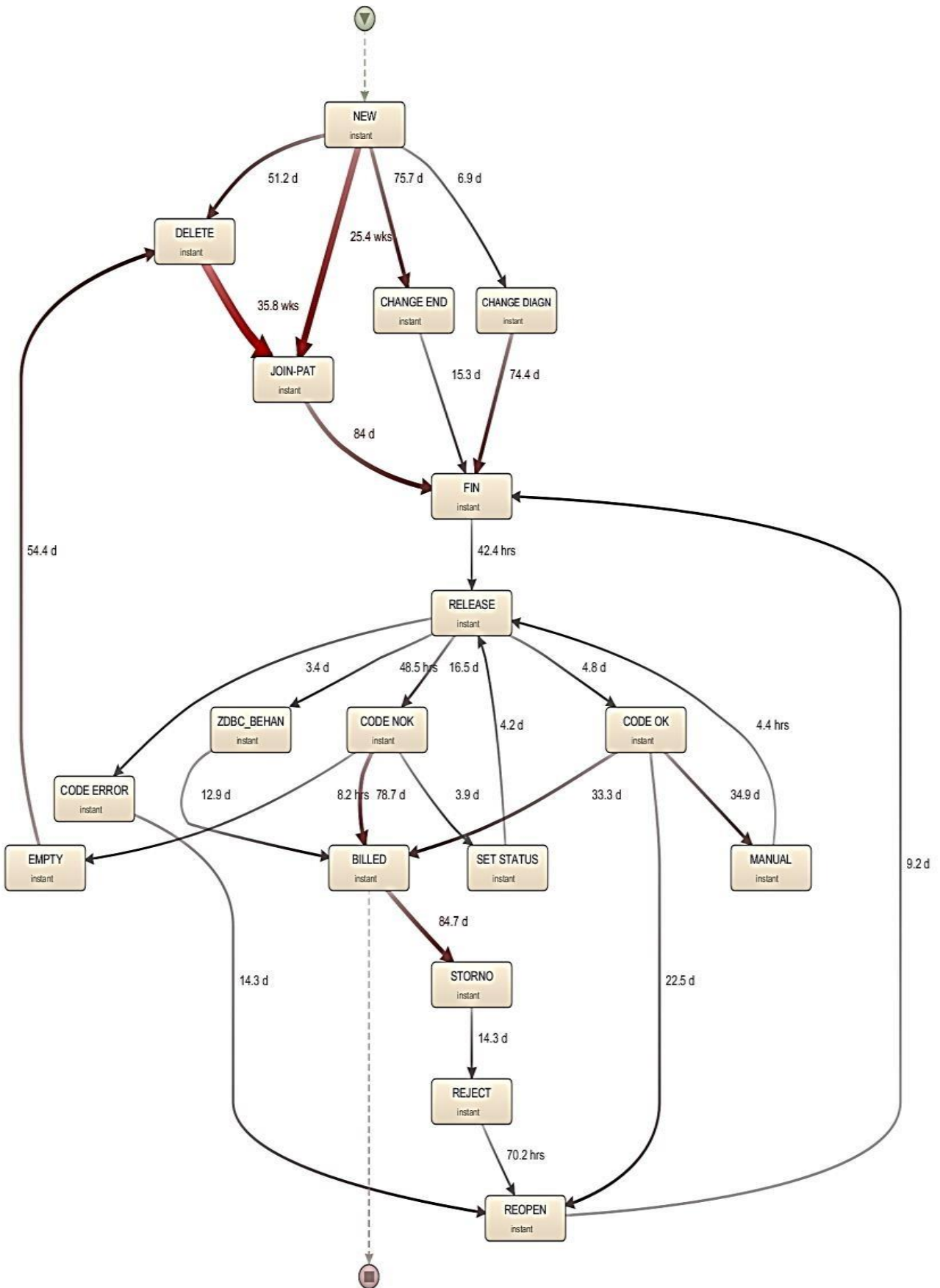
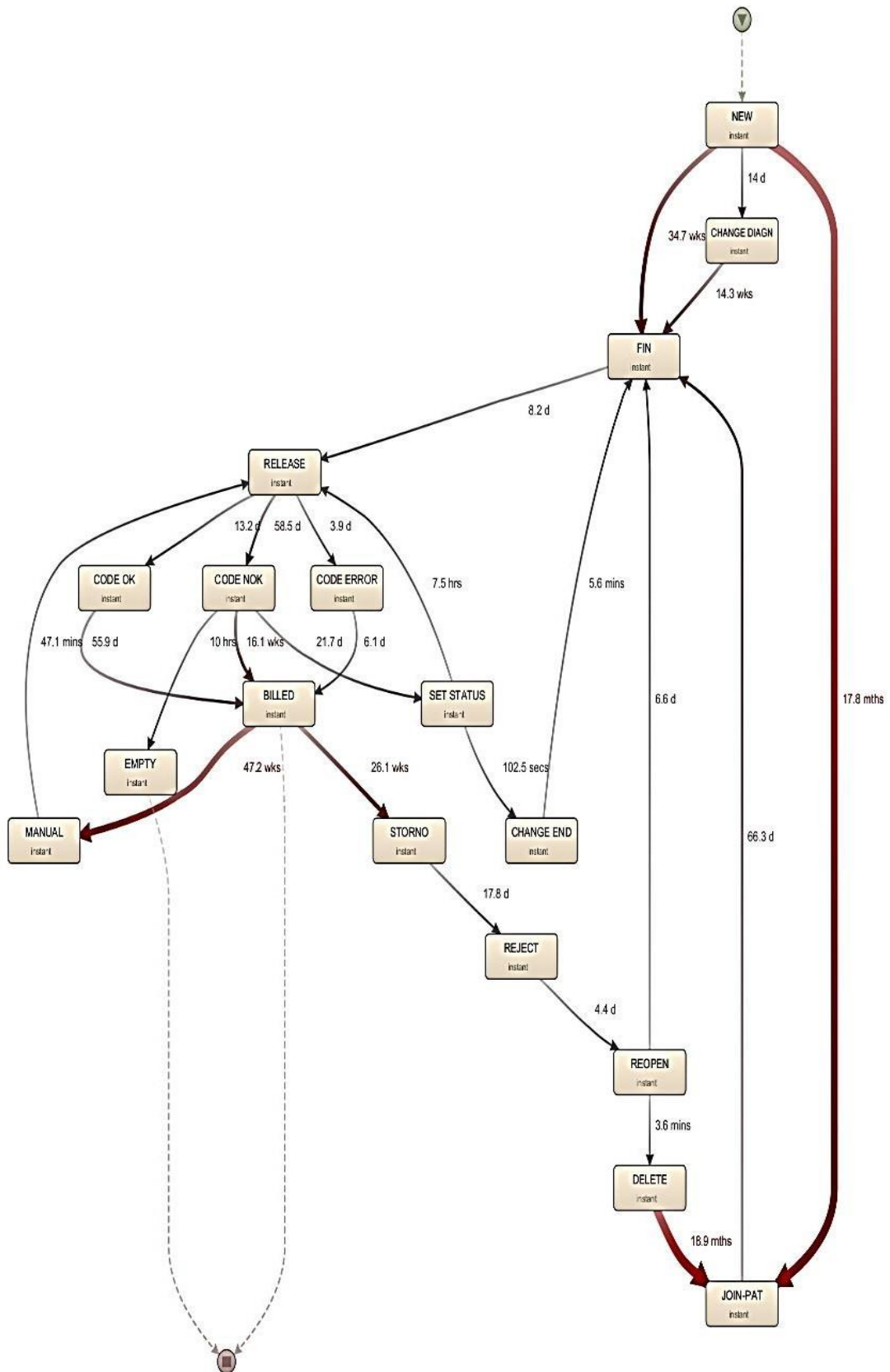*Figure 3 Eventual process model extracted from process discovery.*

Figure 4 Outcome Model by incorporating performance filter techniques

References:

[1] W. M. P. Van Der Aalst and S. Dustdar, "Process mining put into context," IEEE Internet Comput., 2012.

[2] M. Rovani, F. M. Maggi, M. De Leoni, and W. M. P. Van Der Aalst, "Declarative process mining in healthcare," Expert Syst. Appl., 2015.

[3] W. M. P. M. P. van der Aalst, Process Mining: Discovery, Conformance, and Enhancement of Business Processes. 2011.

[4] E. Rojas, J. Munoz-Gama, M. Sepúlveda, and D. Capurro, "Methodological Review Process mining in healthcare: A literature review," J.Biomed. Inform., 2016.

[5] I. Kadi, A. Idri, and J. L. Fernandez-Aleman, "Knowledge discovery in cardiology: A systematic literature review," International Journal of Medical Informatics. 2017.

[6] W. Van Der Aalst, T. Weijters, and L. Maruster, "Workflow mining: Discovering process models from event logs," IEEE Trans. Knowl.Data Eng., 2004.

[7] D. Chapela-Campa, M. Mucientes, and M. Lama, "Mining frequent patterns in process models," Inf. Sci. (Ny)., vol. 472, pp. 235–257, 2019.

[8] Á. Rebuge and D. R. Ferreira, "Business process analysis in healthcare environments: A methodology based on process mining," Inf. Syst.,2012.

[9] R. Gerhardt, J. F. Valiati, and J. V. Canto dos Santos, "An Investigation to Identify Factors that Lead to Delay in Healthcar e Reimbursement Process: A Brazilian case," Big Data Res., vol. 13, pp. 11–20, 2018.

[10] R. Mans, H. Schonenberg, G. Leonardi, S. Panzarasa, A. Cavallini, S. Quaglini, and W. Van Der Aalst, "Process mining techniques: an application to stroke care," Stud. Health Technol. Inform., 2008.

[11] U. Kaymak, R. Mans, T. Van De Steeg, and M. Dierks, "On process mining in health care," in Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics, 2012.

[12] W. M. P. van der Aalst, H. A. Reijers, A. J. M. M. Weijters, B. F. van Dongen, A. K. Alves de Medeiros, M. Song, and H. M. W. Verbeek, "Business process mining: An industrial application," Inf. Syst., 2007.

[13] M. K. M.V., T. L., and A. B., "Capturing the sudden concept drift in process mining," CEUR Workshop Proc., vol. 1371, no. January, pp. 132–143, 2015.

[14] 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," in Communications in Computer and Information Science, 2008.

[15] W. M. P. Van der Aalst and A. J. M. M. Weijters, "Process mining: A research agenda," Comput. Ind., 2004.

[16] A. P. Kurniati, G. Hall, D. Hogg, and O. Johnson, "Process mining in oncology using the MIMIC-III dataset," in Journal of Physics:
Conference Series, 2018.

[17] A. K. A. De Medeiros, A. J. M. M. Weijters, and W. M. P. Van Der Aalst, "Genetic process mining: An experimental evaluation," Data
Min. Knowl. Discov., 2007.

[18] "4TU - Collection: Real life Event Logs." [Online]. Available: https://data.4tu.nl/repository/collection:event_logs_real. [Accessed: 30-Oct- 2018].