Process Mining on BPJS Kesehatan Data Sample for Disease Trajectory Analysis with Secondary Diagnosis

1st Aulia Rahman Arif Wahyudi
School of Computing
Telkom University
Bandung, Indonesia
auliararif@student.telkomuniversity.ac.id

2nd Angelina Prima Kurniati
School of Computing
Telkom University
Bandung, Indonesia
angelina@telkomuniversity.ac.id

3rd Guntur Prabawa Kusuma
School of Applied Science
Telkom University
Bandung, Indonesia
guntur@telkomuniversity.ac.id

***Abstract*— Disease trajectory, the course of a disease over time, and secondary diagnoses, additional medical conditions that a patient may have in addition to their primary diagnosis, can greatly impact patient outcomes, treatment, and management. This study analyzed the feasibility of disease trajectory analysis with secondary diagnoses using the Indonesia Health Insurance (BPJS Kesehatan) 2015-2018 data sample. The study followed the established Process Mining Project Methodology (PM2). We extract the data set from the BPJS Kesehatan data sample, generate an event log from them, discover the disease trajectory by doing process discovery using Heuristic Miner, assess the discovered model using conformance checking, and evaluate it. By analyzing the data sample of Acute Myocardial Infarction (AMI) patients, similar patterns were identified in the 2,100 cases with secondary diagnoses, which can be used to take proactive measures to prevent or manage these secondary diagnoses and gain a more comprehensive understanding of how patients' health changes over time.**

***Keyword* — process mining, healthcare, disease trajectory, secondary diagnosis**

I. INTRODUCTION

A. Introduction

Disease trajectory refers to the probable pathway of a patient with a given condition may have, taking into account the sequence of diagnoses over time, including the development of symptoms and possible complications [1]. This can include factors such as the severity of symptoms, the rate of progression, and the likelihood of recovery or recurrence. Understanding a disease trajectory is crucial for both diagnosis and treatment, as it provides insight into how a disease progresses, the symptoms it presents, and potential complications that may arise.

In recent years, research on disease trajectory has increasingly relied on data from Electronic Health Records (EHRs) and other sources to better understand the natural history of diseases and identify potential targets for intervention [2]. A real-world healthcare dataset is usually difficult to access due to its sensitivity [3]. But the Indonesian Health Insurance (BPJS Kesehatan) has provided real-world data samples of participants of BPJS Kesehatan from 2015-2018. The dataset covers membership data and the electronic health records service of participants acquired from BPJS Kesehatan data warehouse [4]. One of the data sample sub-set data is Advanced Referral Health Facilities Service Secondary Diagnoses.

Secondary diagnoses refer to additional diagnoses that a patient may have in addition to their primary diagnosis. These additional diagnoses include complications that arose during their hospital stay and comorbidities [5]. Secondary diagnoses can have an impact on the patient's treatment, recovery, and prognosis. They also can affect the cost of the treatment and the length of the hospital stay. According to the CMS ICD-10-CM Official Guidelines for Coding and Reporting [6], secondary diagnoses were judged by clinicians when the 2 diagnosis process happens and are coded as 'additional codes' for the primary diagnosis with no sequencing or time factor applied to it.

Disease trajectory and secondary diagnoses can greatly impact patient outcomes, treatment, and management. By analyzing EHR data from patients with a specific patient group, we can identify patterns that may indicate an increased risk of developing certain secondary diagnoses. By identifying these patterns, we can take proactive measures to prevent or manage these secondary diagnoses. In this paper, we analyze the BPJS Kesehatan data sample for disease trajectory analysis using Process Mining [7], which can be utilized to make disease trajectory based on patients' diagnoses [8], creating a Data-Driven model and in turn able to support an Evidence-Based business process. The literature review revealed that recent studies have shown the potential of process mining in the analysis of disease trajectory [2, 9, 10]. Other studies have also shown that the BPJS Kesehatan data sample is viable to be used in disease trajectory analysis [11]. While some studies, like the ones that have been mention previously,

have explored the use of process mining in healthcare, most have focused on primary diagnoses data. Since a patient have the possibility of having multiple secondary diagnoses on one primary diagnoses, the HER will show multiple diagnoses on the same timestamp. And with no established technique to analyze these secondary diagnoses, there is a gap in the literature regarding the use of secondary diagnoses data in disease trajectory analysis.

In this study, we analyzed the BPJS Kesehatan data samples for the feasibility of disease trajectory analysis with insight of secondary diagnoses, using process mining as our primary approach. This study followed the established process mining project methodology (PM2) [12] and focused on specific diagnoses for our analysis. By analyzing event logs from EHRs, process mining can uncover hidden patterns that connect primary diagnoses with secondary diagnoses. By combining this information with data on disease trajectory and secondary diagnoses, we can gain a more comprehensive understanding of how patients' health changes over time and identify potential risk factors for developing additional medical conditions.

II. THEORITICAL REVIEW

A. Methodology

1. General Methodology

We analyze the disease trajectory using process mining [7] by following an established methodology, precisely the Process Mining Project Methodology (PM2) [12]. The main parts of process mining [13] are process discovery, conformance checking, and process enhancement. In process discovery, we utilized the Heuristic Miner algorithm since it has been implemented in various case studies [14] for its ability to handle large data and noise [15]. In conformance checking, we will use replay fitness, precision, and generalization as the evaluation metric. We also used process mining tools such as ProM as an established framework for academic purposes of process mining projects to help in this study. An overview of the method is presented in Figure 1.

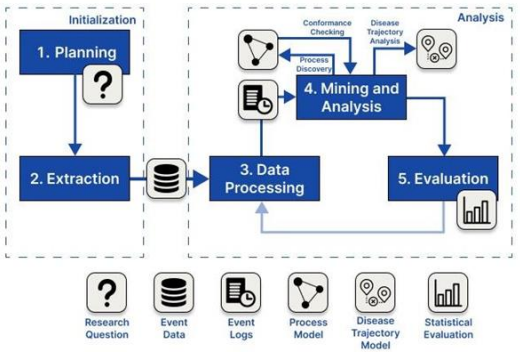

FIGURE 1.
Disease trajectory analysis using process mining methodology overview

Based on Figure 1, the explanation of the research methodology is as follows. The methodology itself follows the Process Mining Project Methodology (PM2) [12], which consists of 5 stages, which are (1) Planning; (2) Extraction; (3) Data Processing (4) Mining & Analysis; and (5) Evaluation.

In the Planning stage of this research, we define the research questions that we will try to answer in this study and the scope of our research, namely a viable diagnosis to be analyzed.

In the Extraction stage, we extract information from the BPJS Kesehatan data sample according to the scope defined in the previous stage. We used the first 3 digits of ICD-10 diagnostic codes to indicate diagnoses, but we excluded the codes known to be not related to the development of diseases (e.g. administration codes). Event data were extracted from FKTPKAPITASI, FKRTL, and SECONDARYDIAGNOSIS tables in the BPJS Kesehatan data sample as the input for generating the event log. The time of visitation was used as the event timestamp and the ICD-10 diagnostic code as the activity itself. Since we will analyze the Secondary Diagnosis, we separated the Primary Diagnosis extracted from the FKTPKAPITASI and FKRTL table apart from the Secondary Diagnosis from the SECONDARYDIAGNOSIS table.

In the Data Processing stage, we plan to generate two event logs from the sample data that has been extracted and preprocessed. We generate it by creating views, aggregating events, enriching data, and filtering logs. The case identifier for each event was taken from the BPJS participant number (CaseID), the ICD-10 diagnostic code as the event activity (Primary Diagnosis and Secondary Diagnosis), and the visitation time as the event timestamp (timestamp). The event log is then filtered by removing recurring diagnoses (keeping the first occurrence) and removing unique traces that is shared by only one case. The sequence of diagnosis codes for each patient in the event log informed the order in which the patient is diagnosed with diseases represented by the ICD-10 diagnostic code, where D1→D2 means that D1 diagnosis preceded D2 diagnosis. The information of which diagnosis preceded which diagnosis is stored on the timestamp of the event, where D1→D2 means that the timestamp for the event with D1 diagnosis occurred earlier than the event with D2 diagnosis. In the case of secondary diagnoses, it is possible for each primary diagnosis to have multiple secondary diagnoses. For example, a patient has a primary diagnosis of N39 has I25, I11, and E11 as their secondary diagnoses.

Since the SECONDARYDIAGNOSIS table in the BPJS Kesehatan data sample doesn't specify the order of the secondary diagnoses diagnosed for the patient, it is difficult to determine the secondary diagnoses pattern (e.g. I25→I11 and I11→I25 will be considered different diagnosis pattern). Thus, for this study, we decided to order the secondary diagnoses alphabetically to form trace variants for the secondary diagnoses and find the patterns formed. As such, following the previous example, then the secondary diagnosis of the patient is E11→I11→I25.

The event log is then imported into the ProM tool. A 'START' and 'END' event was added to every event to provide common start and end points of traces. It is critical to note that the BPJS Kesehatan data sample so far only recorded patients in the period of 2015-2018, which makes it impossible to identify what was the actual first event that a patient was diagnosed since we have no record of what occurred before 2015. The event log is then converted into

XES format and is ready for the next stage of Mining and Analysis.

In the Mining and Analysis stage, we performed process discovery and conformance checking. Process discovery is a technique of taking recorded event logs and learning process models generated from it to get information of real events and user behaviour [13]. In this study, we performed process discovery using Heuristic Miner Algorithm [14]. Heuristic Miner algorithm focuses on the calculation of traces of events and dependency frequency in the process of building an event. The first step of Heuristic Miner algorithm [14] is to make a dependency graph, a matrix based frequency used to show how high is the causal relation between pair of activities. To calculate the value of dependency relation between pair of activities, we can use Equation (1), where $a > Lb$ iff for some case a is directly followed by b.

$$|a \Rightarrow {}_L b| = \begin{cases} \dfrac{|a > {}_L b| - |b > {}_L a|}{|a > {}_L b| + |b > {}_L a| + 1} & \text{if } a \text{ G } b \\ \dfrac{|a > {}_L a|}{|a > {}_L a| + 1} & \text{if } a = b \end{cases} \quad (1)$$

The next step is to build Causal Matrix to represent the correct process model in order to define whether a process is parallel or sequential to another process. For process discovery and the visualization of the discovered model in this study, we used the Interactive Data-aware Heuristic Miner (iDHM) plug-in [16] in the ProM tool. To assess the quality of the discovered process, a conformance checking will be carried out to measure the replay fitness, precision, and generalization [7, 13, 17]. According to [17], replay fitness quantifies the extent to which the model can reproduce traces recorded in the event log. We can calculate the fitness using token-based replay by using Equation (2), where p (produced token), c (consumed token), m (missing token), and r (remaining tokens).

$$\text{Fitness}\,(\sigma, N) = \frac{1}{2}\left(1 - \frac{m}{c}\right) + \frac{1}{2}\left(1 - \frac{r}{p}\right) \quad (2)$$

Precision is a measure of how good the model is at representing traces from the event log. We can calculate the precision of a model by using Equation (3), where n_i (the number of process instances combined) and x_i (the mean number of visible transition).

$$Q_p = 1 - \frac{\sum_{visited\ markings} n_i x_i}{\#total\ marking\ visits\ over\ all\ markings} \quad (3)$$

With x_i calculated in Equation (4).

$$x_i = \frac{\#outgoing\ edges - \#used\ edges}{\#outgoing\ edges} \quad (4)$$

Generalization evaluates how well the generated model will reproduce the future behavior of the process. We can calculate generalization by using Equation (5).

$$Q_g = 1 - \frac{\sum_{nodes}\left(\sqrt{\#executions}\right)^{-1}}{\#nodes\ in\ model} \quad (5)$$

The value of each evaluation metric is computed on a scale from 0 to 1, where 1 is an optimal score. In this study, we used the Multi-perspective Process Explorer plug-in [18] in the ProM tool for measuring the replay fitness and precision, and Measure Precision/Generalization plug-in for measuring the generalization.

In the Evaluation stage, we evaluated the experiment we have done and the results using statistical evaluation and the decision done in this study. We also discussed with an expert who we asked to state their thoughts to confirm our result.

2. Data Sample of BPJS Kesehatan

The Indonesia Social Security Provider (*Badan Penyelenggara Jaminan Sosial* / BPJS) was established as a part of a law by the Government of Indonesia to ensure the basic needs of living for the participants, which include Indonesian citizens and foreigners who had lived in Indonesia for at least 6 months or more [19]. One of the services provided by the BPJS program includes health insurance, which started operation on January 1st, 2014. As the program progresses, countless records of data have been collected, and from these data, we can acquire various information. BPJS Kesehatan has provided a data sample of participants of BPJS Kesehatan from 2015- 2018. The data was previously released in 2019 that covers sample data from the year 2015-2016 and in 2020 they released it again with additional data from the year 2017-2018.

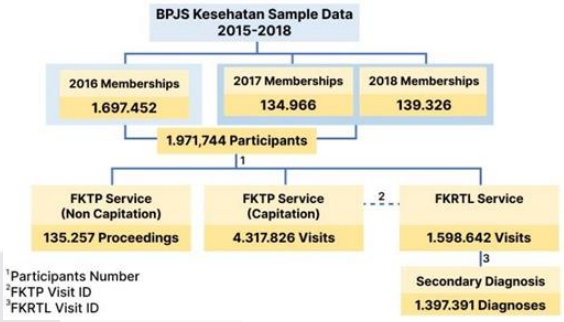

FIGURE 2.
BPJS Kesehatan 2015-2018 data sample hierarchy

The data hierarchy of BPJS Kesehatan data sample 2015-2018 consists of two databases in the form of memberships and health services. Membership data are unique for each participant based on participant ID. While service data Primary Health Facilities (*Fasilitas Kesehatan Tingkat Pertama* / *FKTP*) and Advanced Referral Health Facilities (*Fasilitas Kesehatan Rujukan Tingkat Lanjut* / *FKRTL*) are participants' visits that are submitted to memberships data [20]. The *FKRTL* Service has an extension to the Secondary Diagnosis table, where it shows what secondary diagnoses the patients have along with the primary diagnosis.

III. METHOD

A. Result and Analysis

This study implemented the disease trajectory analysis with secondary diagnosis by following the stages that consisted in the PM^2. The results are presented in these sections following the stages of: (1) Planning; (2) Extraction; (3) Data processing; (4) Mining and Analysis; and (5) Evaluation.

1. Planning

The BPJS Kesehatan data sample has been proven to be viable for process mining projects [21] one of them being disease trajectory analysis [11]. This study aims to analyze the disease trajectory of patients' secondary diagnosis patterns along with the primary diagnoses. For this initial study, we chose a specific disease to be analyzed. For the diagnosis we want to analyze, we selected patients suffering from Acute Myocardial Infarction (AMI) from the BPJS Kesehatan data sample. The diagnosis of AMI patients is indicated by the ICD-10 diagnostic code of I21 (acute myocardial infarction), I22 (subsequent myocardial infarction), and I23 (certain complications following acute myocardial infarction). We checked on the FKRTL Service Secondary Diagnosis table of the BPJS Kesehatan data sample. There are 3,388 patients diagnosed with AMI diseases as either their primary diagnosis or secondary diagnosis. The frequencies of AMI patients are presented in Table 1.

TABLE 1.
Frequencies of patients diagnosed with AMI diseases in the BPJS Kesehatan data sample

ICD-10 Code	Diagnosis Name	# of Occurences	# of Patients
I21	Acute Myocardial Infarction	5,622	3,173
I22	Subsequent Myocardial Infarction	237	138
I23	Certain Complications following Acute Myocardial Infarction	315	143

Studies show that Acute Myocardial Infarction is a leading cause of morbidity and mortality worldwide [22], with many AMI deaths occurring within 28 days of being in a hospital [23]. For those reasons, we continued with AMI patients selected to be analyzed in this study.

Following the literature review, we define the main research questions for this study as follows: Question 1. Can disease trajectories of secondary diagnoses be identified using process mining approach? Question 2. What are the common secondary diagnoses for AMI patients? Question 3. What are the most common primary diagnosis with AMI diagnoses group as their secondary diagnosis?

2. Extraction

There are 6,051,725 unique visits from the 1,971,744 patients in the BPJS Kesehatan data sample, and of those visitations, there are 1,170 unique ICD-10 diagnostic codes. First, we excluded 135,257 (2.2%) visitations from the FKTPNONKAPITASI table since patients from that table doesn't have diagnosis code. We further excluded 920,489 (15.2%) visitations that have diagnosis codes that are not related to the development of diseases [24], this includes codes about pregnancies (Chapter XV and XVI of ICD-10), general symptoms and signs not linked to a disease (chapter XVIII), external causes (chapters XIX and XX) and administration (chapter XXI). We finally excluded visitations for patients that are never diagnosed with AMI diseases as their primary diagnosis or their secondary diagnosis. For this initial study, instead of taking individual diagnoses of AMI patients (i.e. I21, I22, and I23), we represent them as a diagnosis group and transform those diagnostic codes into 'AMI' and take the first occurrence as the start of the AMI diseases on the patient. The final dataset leaves us with 19,958 visitations from 3,388 patients.

3. Data Processing

We generated two event logs from the dataset, first is the event log containing primary diagnoses for the disease trajectory of AMI patients. The second event log contains the alphabetical order of secondary diagnoses from AMI patients. From this point further, we will refer to the first event log as PD (Primary Diagnosis) Event log and the second event log as SD (Secondary Diagnosis) Event Log. We composed the extracted dataset from the previous stage to follow the minimum requirements of an event log. We load the event log into ProM and analyze the event log with basic characteristics presented in Table 2.

Table 2. Characteristics of the event logs

Characteristics	PD Event Log	SD Event Log
Cases	3,388	2,100
Events	19,958	5,529
Event classes	653	294
Events per case	Min: 1; Mean: 6; Max: 44	Min: 1; Mean: 3; Max: 28
Event classes per case	Min: 1; Mean: 6; Max: 44	Min: 1; Mean: 2; Max: 18
Variants	2,616	1,226
Event per trace	5,891	2,633

After that, we added artificial 'START' and 'END' events and did some filtering to event logs, which is merging subsequent events by removing the recurring diagnoses for each patient and keeping the first occurrences, and filtered out traces that are unique (only shared by one case).

IV. RESULTS AND DISCUSSION

A. Mining & Analysis

The mining activity consists of process discovery and conformance checking. For process discovery, we used Heuristic Miner for its ability to handle large data and noise. In this stage, we only included multiple traces of the event logs by filtering out unique traces. The final event log for PD Event Log consists of 867 cases with 3,113 events of 39 event classes, with 93 trace variants. We put the event log into the ProM mining tool and applied the Interactive Data-aware Heuristic Miner (iDHM) plugin.

The discovered model for the PD Event Log is presented as a Directly-follows Graph in Figure 3.

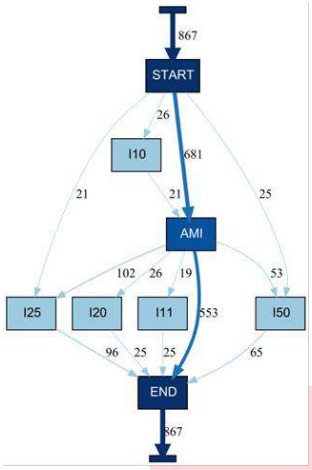

FIGURE 3.
The Directly-follows Graph representation of disease trajectory model of ami patients in BPJS Kesehatan data sample with minimum case frequency = 18

Figure 3 shows that there is one diagnosis that commonly comes before the diagnosis of AMI diseases, namely essential (primary) hypertension (I10). Meanwhile, there are five diagnoses that commonly come after the diagnosis of AMI disease, namely hypertensive heart disease (I11), angina pectoris (I20), chronic ischaemic heart disease (I25), and heart failure (I50). The conformance checking for the discovered model demonstrated trace fitness = 0.948, precision = 0.936, and generalization = 0,972. The numbers show that the discovered process model has been representative of the records in the PD Event Log.

On the other hand, the final event log SD Event Log consists of 1,020 cases with 3,328 events of 68 event classes, with 146 trace variants. The event log is put into the ProM mining tool and applied the iDHM plugin, and the discovered model for the SD Event Log is presented as a directly-follows Graph in Figure 4.

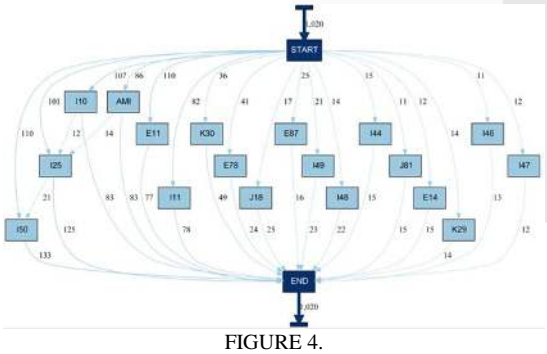

FIGURE 4.
The Directly-follows Graph representation of disease trajectory model with secondary diagnosis from AMI patients in BPJS Kesehatan data sample with the minimum case frequency = 11

The Directly-follow Graph shows nodes that correspond to activities and directed edges that correspond to directly-follows relationships. Note that The DFG presented in Figure 3 is distinct with the DFG presented in Figure 4. Figure 4 shows the trajectory of secondary diagnoses from AMI Patients. However, the model was only discovered to help identify trace variants of secondary diagnoses, not the progressiveness of secondary diagnoses. We can interpret the model as "the variants of AMI patients secondary diagnoses". We can see that most trajectory only consists of one secondary diagnosis, but there are trajectories that show a course of secondary diagnoses. For example, we can see that there is a trajectory of START→I10→I25→I50→END. This means that as the AMI patient is being diagnosed, the secondary diagnoses appear in this pattern.

The conformance checking of the discovered model demonstrated trace fitness = 0.923, precision = 0.993, and generalization = 0.979. The numbers show that the discovered process model has been representative of the records in the SD Event Log. In the SD Event Log, there were 146 unique trace variants of secondary diagnoses of AMI patients that informed the process discovery algorithm to answer Question 1, "Can disease trajectories of secondary diagnoses be identified using process mining approach?".

In response to Question 2, "What are the common secondary diagnoses for AMI patients?", among the 1,020 patients, there are 146 distinct trajectories. Figure 5a shows that the most common (84 Cases; 6.25%) secondary diagnosis for AMI patients is heart failure (I50). For trace variants with multiple secondary diagnoses, Figure 5b, and Figure 5c show that chronic ischaemic heart disease (I25) and heart failure (I50) as the most common (15 Cases; 2.59%) secondary diagnoses with two diagnoses; and other disorders of fluid, electrolyte and acid-base balance (E87), hypertensive heart disease (I11), and other cardiac arrhythmias (I49) as the most common (3 Cases; 0.91%) secondary diagnoses with three diagnoses.

The third question was "What is the most common primary diagnosis with AMI diagnoses group as their secondary diagnosis?". The top three most common primary diagnoses are chronic ischaemic heart disease (I25) with 34 events (27.2%), heart failure with 18 events (14.3%), and hypertensive heart disease (I11) with 8 events (6.4%).

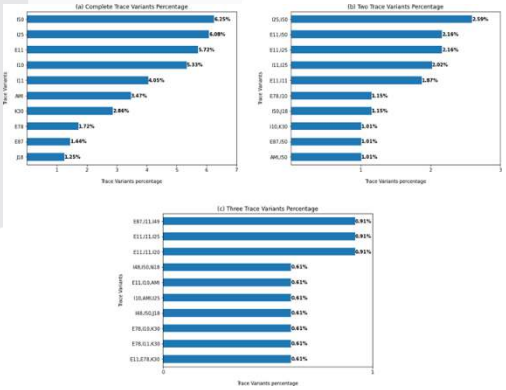

FIGURE 5.
Top 10 Secondary Diagnoses Trace Variants

B. Evaluation

Our study has shown the possibility of process mining for disease trajectory with secondary analysis. We followed the PM2 as our main methodology and perform process discovery using Heuristic Miner algorithm. We evaluate how we chose our selection of cohort patients, where we

analyzed the data in the BPJS Kesehatan data sample and chose a specific diagnosis based on the insight we got. We chose patients diagnosed with acute myocardial infarction (AMI) consisting of 3,388 patients. This found a number of patients having similar primary diagnosis disease trajectories. Furthermore, we made a second event log to analyze the secondary diagnoses of AMI patients. Out of the 2,100 cases that have secondary diagnoses, we found patterns of diagnoses that were shared by multiple patients. We apply the Interactive Data-aware Heuristic Miner (iDHM) plugin in the ProM tool for our process discovery, and we used Directly-follows Graph (DFG) to present the discovered model because of its simple presentation that can help in our understanding of the model. However, these DFGs may be misleading and readers need to know how these process models are generated before interpreting them [25]. For conformance checking, we used measure replay fitness, precision, and generalization as the three metrics that reflects the representativeness of the discovered model. Both discovered models for primary diagnosis and secondary diagnosis demonstrated fitness, precision, and generalization scores above 0.9, and with a scale from 0 to 1 -with 1 as an optimal score- the discovered models has been representative of the records in both event logs . A clinician expert, dr. Yudi Wahyudi, SpPD-KGEH from Santosa Hospital was asked to testify their opinion on the result of our experiment. The expert confirms that the result of our experiment is consistent with what he have seen in their experience as a clinician. However, he further stated that since the data set is secondary data (i.e. collected by someone else earlier, in this case, the BPJS Kesehatan), the data tends to lack accuracy and quality to our research. To which he suggest carrying out future research with primary data to improve the accuracy and quality of the result.

V. CONCLUSION

A. Conclusion

This study has shown the feasibility of process mining for disease trajectory analysis with secondary diagnosis. The mining was conducted using the EHR we extracted from the BPJS Kesehatan data sample. By analyzing the EHR, patterns were discovered between primary diagnosis and secondary with the chosen patient group. This information can be used to take proactive measures to prevent or manage these secondary diagnoses and gain a more comprehensive understanding of how patients' health changes over time. The study followed the established PM^2 and used the iDHM plugin in ProM as our main algorithm in process discovery. Future works of this study can further develop the methodology in this study, utilize the technique used in this study to explore the secondary diagnoses in other EHR, and/or use a much more updated version of the BPJS Kesehatan data sample which is expected to be more detailed and populated.

REFERENSI

[1] Larvin, H, Kang, J, Aggarwal, V R, Pavitt, S, and Wu, J. 2021. Multimorbid disease trajectories for people with periodontitis. Journal of Clinical Periodontology, 48(12): 1587–1596.

[2] Kusuma, G P, Kurniati, A P, Rojas, E, McInerney, C D, Gale, C P, and Johnson, O A. 2021. Process mining of disease trajectories: A literature review. Studies in Health Technology and Informatics.

[3] Munoz-Gama, J, Martin, N, Fernandez-Llatas, C, Johnson, O A, Sepúlveda, M, Helm, E, Galvez-Yanjari, V, Rojas, E, Martinez-Millana, A, Aloini, D, Amantea, I A, Andrews, R, Arias, M, Beerepoot, I, Benevento, E, Burattin, A, Capurro, D, Carmona, J, Comuzzi, M, Dalmas, B, de la Fuente, R, Di Francescomarino, C, Di Ciccio, C, Gatta, R, Ghidini, C, Gonzalez-Lopez, F, Ibanez-Sanchez, G, Klasky, H B, Kurniati, A P, Lu, X, Mannhardt, F, Mans, R, Marcos, M, Medeiros de Carvalho, R, Pegoraro, M, Poon, S K, Pufahl, L, Reijers, H A, Remy, S, Rinderle-Ma, S, Sacchi, L, Seoane, F, Song, M, Stefanini, A, Sulis, E, ter Hofstede, A H M, Toussaint, P J, Traver, V, Valero-Ramon, Z, Weerd, I van, van der Aalst, W M P, Vanwersch, R, Weske, M, Wynn, M T, and Zerbato, F. 2022. Process mining for healthcare: Characteristics and challenges. Journal of Biomedical Informatics, 127: 103994.

[4] 2013. Sejarah Perjalanan Jaminan Sosial di Indonesia. [Online] Available at: https://bpjs-kesehatan.go.id/bpjs/index.php/pages/detail/2013/4; [Accessed 24 Jan. 2023].

[5] Iezzoni, L I. 1992. Comorbidities, complications, and coding bias. does the number of diagnosis codes matter in predicting in-hospital mortality?. JAMA: The Journal of the American Medical Association, 267(16): 2197–2203.

[6] 2022. ICD-10-CM Official Guidelines for Coding and Reporting. [Online] Available at: https://www.cms.gov/files/document/fy-2022-icd-10-cm-coding-guidelines-updated-02012022.pdf [Accessed 24 January 2023]

[7] van der Aalst, W M. 2018. Process Mining Data Science in Action. Springer.

[8] Fernandez-Llatas, C. 2021. Interactive process mining in Healthcare. Cham, Switzerland: Springer.

[9] Kusuma, G, Sykes, S, McInerney, C, and Johnson, O. 2020. Process mining of disease trajectories: A feasibility study. Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies.

[10] Kusuma, G, Kurniati, A, McInerney, C D, Hall, M, Gale, C P, and Johnson, O. 2021. Process mining of disease trajectories in Mimic-III: A case study. Lecture Notes in Business Information Processing: 305–316.

[11] Kurniati, A P, Kusuma, G P, and Wisudiawan, G A. 2022. Process mining for disease trajectory analysis on the Indonesia Health Insurance Data. JURIKOM (Jurnal Riset Komputer), 9(5): 1322.

[12] van Eck, M L, Lu, X, Leemans, S J, and van der Aalst, W M. 2015. PM^2: A process mining project methodology. Advanced Information Systems Engineering: 297–313.

[13] Garcia, C dos, Meincheim, A, Faria Junior, E R, Dallagassa, M R, Sato, D M, Carvalho, D R, Santos, E A, and Scalabrin, E E. 2019. Process mining techniques and applications – A systematic mapping study. Expert Systems with Applications, 133: 260–295.

[14] Kurniati, A P, Kusuma, G P, and Wisudiawan, G A. 2016. Implementing heuristic miner for different types of event logs. International Journal of Applied Engineering Research 11.8: 5523-5529

[15] Caesarita, Y, Sarno, R, and Sungkono, K R, 2017. Identifying bottlenecks and fraud of business process using Alpha ++ and heuristic miner algorithms (case study: CV. Wicaksana Artha). 2017 11th International Conference on Information & Communication Technology and System (ICTS).

[16] Mannhardt, F, De Leoni, M, and Reijers, H A. 2017. Heuristic Mining Revamped: An Interactive, Data- aware, and Conformance-aware Miner. In Bpm (demos).

[17] Buijs, J C, van Dongen, B F, and van der Aalst, W M. 2012. On the role of fitness, precision, generalization and simplicity in process discovery. On the Move to Meaningful Internet Systems: OTM 2012: 305–322.

[18] Mannhardt, F, De Leoni, M, and Reijers, H A. 2015. The multi-perspective process explorer. BPM (Demos), 1418: 130-134.

[19] Rolindrawan, D. 2015. The impact of BPJS health implementation for the poor and near poor on the use of Health Facility. Procedia - Social and Behavioral Sciences, 211: 550–559.

[20] Abdullah, A A. 2020. Makalah Data Sampel BPJS Kesehatan. BPJS Kesehatan

[21] Kurniati, A P, Wisudiawan, G A, and Kusuma, G P. 2022. Potentials of clinical pathway analysis using process mining on the Indonesia National Health Insurance Data Samples: An Exploratory Data Analysis. 2022 International Conference on Data Science and Its Applications (ICoDSA).

[22] Reed, G W, Rossi, J E, and Cannon, C P. 2017. Acute myocardial infarction. The Lancet, 389(10065): 197– 210.

[23] Asaria, P, Elliott, P, Douglass, M, Obermeyer, Z, Soljak, M, Majeed, A, and Ezzati, M. 2017. Acute myocardial infarction hospital admissions and deaths in England: A national follow-back and follow-forward record-linkage study. The Lancet Public Health, 2(4).

[24] Jensen, A B, Moseley, P L, Oprea, T I, Ellesøe, S G, Eriksson, R, Schmock, H, Jensen, P B, Jensen, L J, and Brunak, S. 2014. Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients. Nature Communications, 5(1).

[25] van der Aalst, W M P. 2019. A practitioner's guide to process mining: Limitations of the directly-follows graph. Procedia Computer Science, 164: 321–328.