Mining Hospital Processes: A Case Study

Chae-Young Kim[1], Ha-Na Kang[2], Hyun-Seok Hwang[3**]

Abstract--- *The growing demand for medical services is raising in healthcare. In addition, due to the development of data information systems, all information related to the medical care process in hospitals is stored and attempts to analyze it are increasing. Accordingly, this study aims to comprehend the medical care process by utilizing process mining technique. The data was collected over three months, and the data consist of information on the medical care process of patients. To identify the overall process, the data was classified by age, gender, medical department and payment type. Contrary to expectations that there would be differences between the two activities (Test Registration, Tests) by classification, the difference was not significant. However, under the activities 'general payment' and 'unmanned payment', there was significant difference. Through the analysis, this study provides more efficient payment type was identified in the medical care processes.*

Keywords--- *process mining, hospital process, bupaR, data mining*

I. Introduction

The environment that surrounds the medical sector of Korea is facing a significant change due to the recent increase in demand for medical services. The domestic medical expense per capita is in continuous increase, and the competition of the medical market is becoming fierce as tertiary care institutions expand, specialization of examination parts and hospitals [1]. Unlike the past practice that recorded all patient information by hand, the current practice of hospital uses electronic records and saves all information generating within the hospital on the hospital management system thanks to the development and implementation of the hospital information system [2]. The saved data act as that data in comprehending the overall treatment process and its problems, such as on what process that the patients are treated or on which section to bottleneck occurs [3]. The arrival of the era of healthcare Big Data is demanding cost reduction and medical service improvement using the massive amount of data, and it is becoming a major challenge for medical institutions [4]. Therefore, the domestic medical institutions of Korea seek to find the method that involves scientific analysis on the medical process for cost reduction and efficiency improvement, and such attempts are increasing [5]. In this context, this thesis seeks to deduct the process model based on the hospital data and comprehend the hospital treatment process by analyzing the flow of the data.

Chae-Young Kim[1], Ha-Na Kang[2], Graduate School of Interaction Design, Hallym University South Korea
*[*]*Corresponding Author: Hyun-Seok Hwang[3], Department of Business Administration, Hallym University, South Korea.*
E-mail: [1]vnfmsquf55@gmail.com, [3]khnnn060@naver.com, [3]hshwang@hallym.ac.kr

II. Related Works

2.1 Process Mining

Process mining is one method of data mining, a technology that helps you discover new pattern and linkage of data [6], and it refers to the process of mining out the meaningful knowledge from event logs that generate along with the process execution within a certain system [7]. This technology aims to detect and improve the issue so that the user buys the product more efficiently, and employees can handle work more efficiently [8,9]. The process exists in all sectors that have a beginning and ending phases, and process mining apply to all processes. The process mining helps the user to find out which activity is the bottleneck within a certain process, which part of the process does the rework occur, and whether if there is an abnormal process deviation, allowing the user to find the room for improvement based on the findings.

The process mining requires data in the event log form for the application. The elements that the event log must include are “case ID,” “Activity,” and “Timestamp.” “case ID” is a unique value that classifies certain clients or patients. “Activity” refers to the activities that a client performs, such as order or purchase, and the activities that a patient performs, such as treatment and checkout. “Timestamp” refers to the time in which the activity has occurred. On top of these types of information, the analysis may include columns, such as gender, age, and cost.

Many sectors, including production, healthcare, logistics, and spectator movement analysis, are implementing the process mining. Especially, the main objective of the process mining in the hospital sector is to identify and improve the waiting/delay time and mistake pattern per examination or medical department. ProM and Disco are the representative process mining tools, and they are in use at various industries and companies [10].

2.2 Hospital Case Study

Process mining can prove useful in overcoming various issues or future challenges of medical institutions around the world. Especially, the aging, the serious world-wide problem, leads to the rapid increase of the medical expense, and many organizations in the world are seeking to reduce the cost using process mining [11].

Mans, Van der Aalst & Vanwersch (2015) have proposed the reference model that is potentially utilizable in the medical sector, and they provided the index for a smoother utilization of process mining by medical institutions [4]. Additionally, AMC, the Dutch University hospital, has deducted the standard process through process mining, and it expanded the process to the entire company to accomplish guideline proposal, reinforcement, and reduction of waiting and consultation time [12]. Cho, Song and Yoo. (2013) have established the simulation model that optimizes the appointment reservation schedule [3]. Choi, Kim & Lee (2015) have utilized emergency room data to propose the model that identifies and predicts the treatment process according to the characteristic information of patients [11]. Lee, Lee and Cha (2017) proposed IoT-based data gateway development technology to effectively manage the energy of hospital facilities [13]. Jung, Kim and Kang (2016) predicted discharge rate of patient using data mining and decision tree [14]. Also, Zhang and Wan (2019) focused on investigating and identifying the durative persuasive design characteristics of healthcare related mobile application through a data mining process [15].

In general, the medical process is very dynamic, intricate, and temporary. Therefore, the system improvement through the process mining in the medical sector bears its limits. However, many medical institutions utilize process mining to analyze and improve the medical process successfully [5].

Thereupon, we propose the following research questions before we analyze the hospital processes.

RQ1: Will there be an age-specific difference in examination-related activities?

RQ2: Will there be a difference in the treatment process due to the gender?

RQ3: Will there be a difference in the treatment process due to the medical department?

III. RESEARCH FRAMEWORKS

For the hospital process analysis, this research has developed the study in the order depicted in "Figure 1." First, the collected data is pre-processed to transform into the form applicable for process mining. Secondly, the process map is deducted, and the analysis takes place. The process analysis is conducted separately for age, gender, medical department, and payment type. In the last phase, the analysis result is interpreted and discussed.

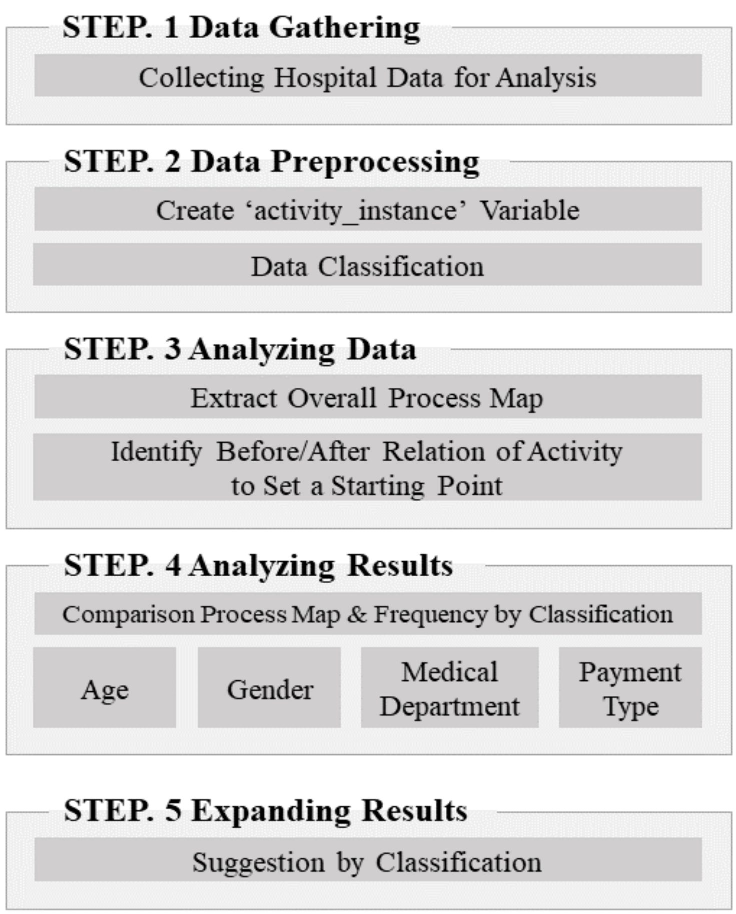

Figure 1. Research Procedure

For the data analysis, the bupaR packaged provided by "R," the open-source analysis tool. The major functions of bupaR is to load event data, explanation generation, process monitoring, and visualization, and its strength is that it supports many phases that analyze the process. "bupaR" consists of eight types of packages, and "bupaR" package is the core package which enables the user to establish the framework. There are six types of information required for the process analysis using the bupaR: case identification, timestamp, activity identification, activity instance ID, resource ID, and lifecycle ID.

IV. Case Study

This chapter covers the hospital process case study based on the framework proposed in previous Chapter. First, the hospital treatment data is selected, and the characteristics of the factor were analyzed.

The data used in the study records all process that patient experiences, from the admission to checkout. By analyzing the hospital data with inconsistent patterns and fluctuations, the research anticipates that it will obtain some understandings of characteristics of the hospital process and the process itself.

4.1 Data Gathering

The treatment data for three months of operation from a large hospital located in Seoul, Korea, starting from June 2018 to September 2018, was utilized for the analysis. The number of data is 108,321 in total, and the information consists of the treatment process information of 9,897 patients. The dataset includes factors, such as “case,” “activity,” and “abs_activity.” Among them, “case” is in the form of “Patient_1000***,” referring to the ID that identifies patients. “Activity” consists of nineteen treatment-related actions, and it is the sub-category that consists of “abs_activity.” “abs_activity” is the superordinate concept, and it consists of eight sub-categories: arrival, treatment, test registration, tests, unmanned payment, general payment, therapy, and walk-in registration.

4.2 Data Preprocessing

The data was pre-processed before the analysis. Since the provided data lacks the factors related to the activity instance, the essential factors for bupaR, as mentioned in Chapter 3, the new value was generated using Excel based on abs-activity and activity, and the factor name of the original data was changed. “Table 1” shows the explanation of the changed factor name and explanations on factors.

Table 1. ***The explanation of the changed factor name and explanations on factors.***

Factor Name	Changed Factor Name	Explanation
case	case	A case identifier
activity	status	A transactional life cycle stage
abs_activity	activity	An activity label
resource	resource	A resource identifier
timestamp	timestamp	A time stamp
-	activity_instance	An activity instance identifier
age	age	-
gender	gender	-

After the factor names were changed, the data were analyzed according to the gender/age/medical department/payment type, depending on the objective of the analysis. The age-specific data were divided into three types (20s to 30s/40s to 50s/60s to 70s), and the medical department data were divided into two types (internal medicine department/surgery department).

International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 7, 2020
ISSN: 1475-7192

4.3 Analyzing Data

The process map was proposed using the total dataset to confirm the overall process of the hospital. The general process map is as shown in "Figure 2."

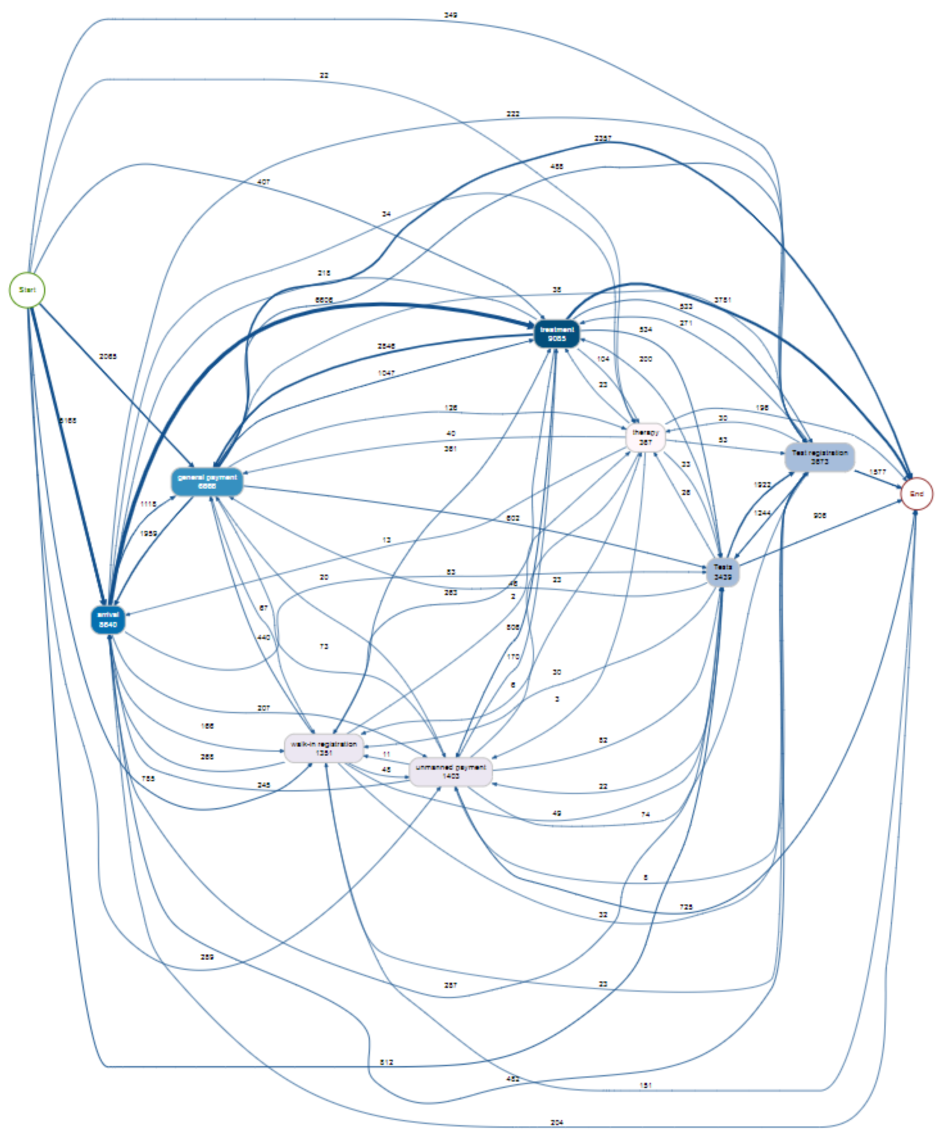

Figure 2. The general process map

The general process map shows a very intricate and unstructured form as in "Figure 2." Therefore, the data, target of the analysis, need to be filtered to simplify and conduct analysis on the process map. This research sought to understand the precedence relations between activities that occur relatively frequently to identify the start point, and it used the "precedence_matrix" function of bupaR. The "Figure 3" is the plot deducted using the "precedence_matrix" function.

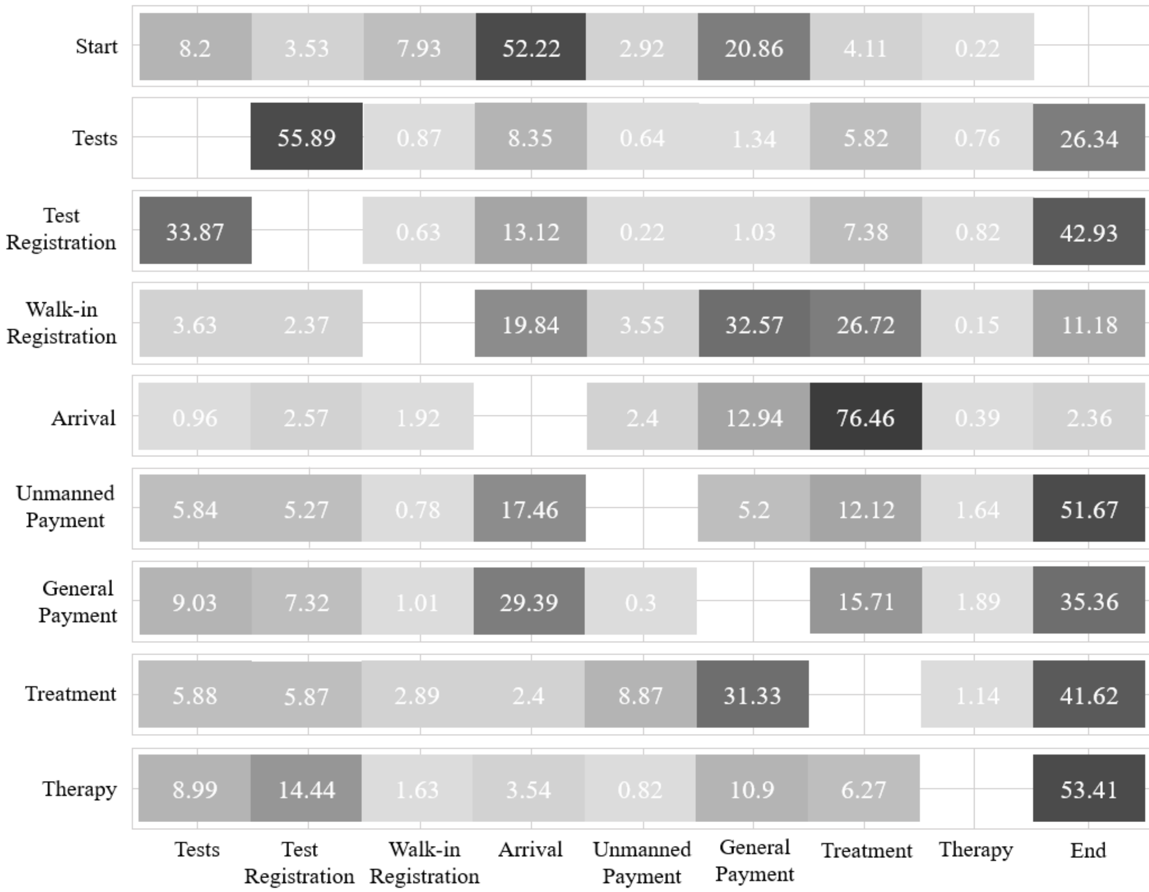

Figure 3. The plot to identify the start point

The ratio of data starting with “arrival” was 52.22%, showing the highest correlation with the start point. Therefore, the start point of all process analysis was set to “arrival.”

In the age-specific analysis, the difference between “tests” and “test registration” was confirmed by comparing the frequency, ration, and the process map using the dataset divided into three types. The process_map() function of bupaR was used for the deduction of the process map. Here, including the trace with relatively too low-frequency cause difficulty in understanding the treatment process, and only the trace of the top 80% was included in the analysis.

As the next step, the process map and throughput time of each activity were compared to identify the difference in the treatment process between males and females. The process map includes the top 70% of the total trace, and the relative frequency was marked. Additionally, the “throughput_time()” function was utilized to compare the throughput time, and the unit was set to “by the minute.” The process map, number of cases per age, and frequency were compared to identify the difference between process per medical department. The top 80% of the total tace was made to appear on the process map.

IV. RESULTS

This chapter explains the comparison result of the frequency and process map per gender/age/medical department/payment method that the preceding chapter has analyzed.

5.1 Analysis Results of Age

In the age-specific data, the "test registration" and "tests" were expected to show a difference, but it turned out that the difference is insignificant. The frequency and ratio of the activity are as shown in "Table 2."

Table 2. ***The frequency and ratio of the age-specific activity***

Age	F	Test Registration	Tests	General Payment	Unmanned Payment
20-30	A	591	556	931	297
	R	0.112	0.105	0.176	0.056
40-50	A	1088	1020	1921	419
	R	0.106	0.099	0.188	0.041
60-70	A	1482	1381	2894	424
	R	0.105	0.097	0.204	0.030

*F: Frequency A: Absolute Frequency, R: Relative Frequency

However, the "general payment" and "unmanned payment" showed a difference. In the 20s to 30s age group, the frequency of the "general payment" was 0.176, and the frequency of the "unmanned payment" was 0.056. In the 40s to 50s age group, the frequency of the "general payment" was 0.188, and the frequency of the "unmanned payment" was 0.041. In the 50s to 60s age group, the frequency of the "general payment" was 0.204, and the frequency of the "unmanned payment" was 0.029; the higher the age, the frequency of using the unmanned payment method was lower.

The age group of the 40s to 50s did not show a significant difference compared to the other data within the process map, and the data from the group was excluded from the process analysis. The ratio of moving on from "tests" to "test registration was 65% for the 20s to 30s age group, and the same ratio for the 60s to 70s age group was 47.62%. In contrast, the ratio of moving on from "test registration" to "tests" was 35% for the 20s to 30s age group, and the same ratio for the 60s to 70s age group was 52.38%. Considering that the typical treatment system of hospital proceeds from test registration to the test, the data indicates that the number of tests is higher as the age decreased. Additionally, in the data of the 20s to 30s age group, the "treatment" activity is included in the process map, but the activity is not included in the process map of the 60s to 70s age group, indicating that the group with a lower age undergoes a relatively various activities within the treatment process compared to the group with a higher age.

5.2 Analysis Results of Gender

The comparison of the frequency, ratio, and process maps of the occurred activities did not show a significant difference between the two gender groups. Similarly, despite the similarity of the process map composition and the number of cases per age group, the only difference was found in the throughput time. The throughput time for the female group was 93.68 minutes, and the time for the male group was 99.63 minutes; the female group spent six minutes less compared to the male group. "Table 3" shows the throughput time of the two groups, and "Figure 4" and "Figure 5" show the process map of a male/female group with relative frequency.

Table 3. ***Throughput time of the two groups***

Gender	Min	Max	Mean	St_dev
Female	0.00	747.00	93.68	93.15
Male	0.00	907.00	99.63	96.05

*unit: mins

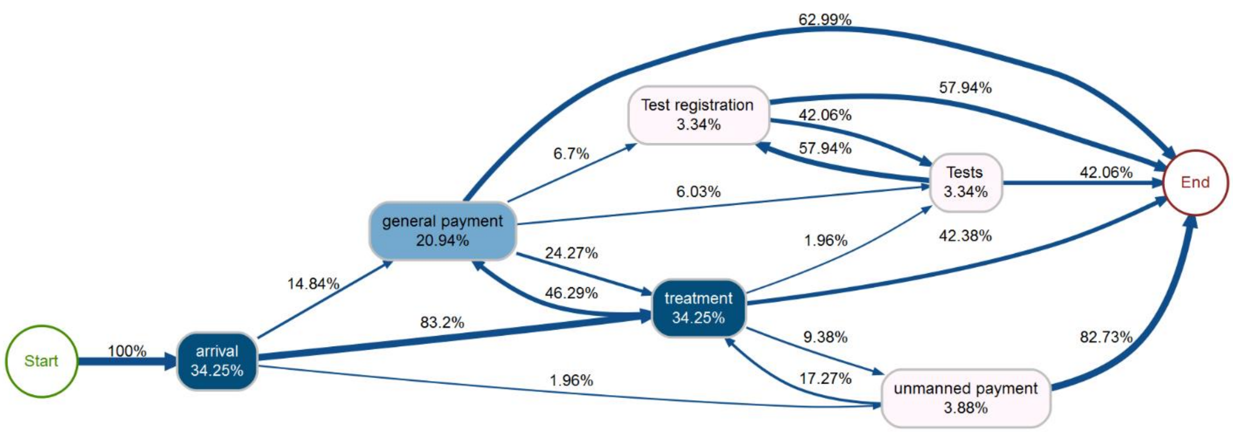

Figure 4. The process map of female group

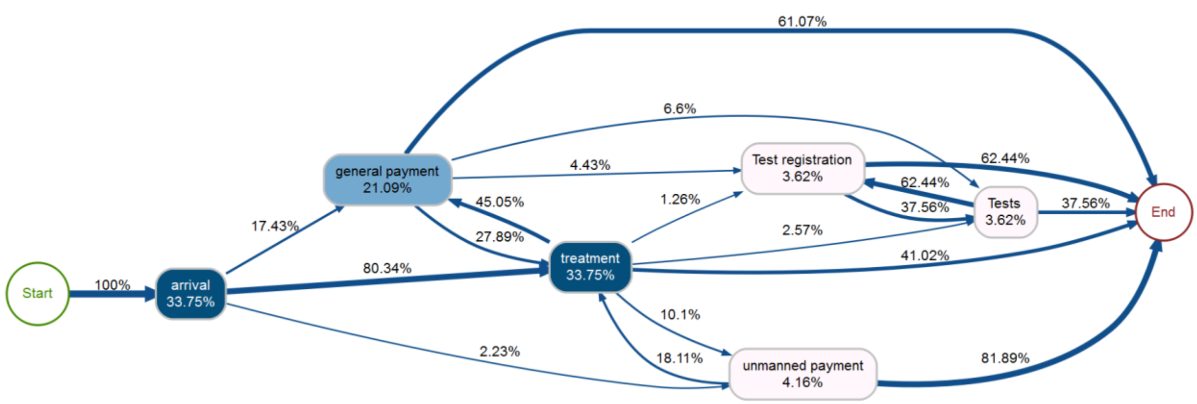

Figure 5. The process map of male group

Additionally, the 57.94% of the female group and the 62.44% of the male group showed the movement starting from "tests" to "test registration," indicating that the male group showed a higher ratio of making another registration after finishing a test; this is due to the tendency of the male group that undergoes more tests compared to the female group.

5.3 Analysis Results of Mecial Department

The analysis result per medical department did not show a significant difference, unlike the initial assumption that the frequency of "test registration" and "tests" would show a significant difference. However, the number of cases per age group of surgery/internal medicine departments shows that the ratio of visiting the internal medicine department rose drastically as the age increased. "Table 4" shows the number of cases per age group and medical department.

Table 4. ***The number of cases per age group and medical department***

Department \ Age	20	30	40	50	60	70
Internal Medicine	169	266	446	798	1,047	967
General Surgery	124	107	189	346	390	290

As the next step, the research examined the process maps for two departments, and it found that the "treatment" activity was marked only on the process map of the surgery department. The result indicates that the "treatment" is not an important activity within the process of the internal medicine department, whereas it is an activity with considerable importance within the treatment process of the surgery department. Additionally, the ratio of movement that starts from "tests" to "test registration" was 43.48% for the internal medicine department and 34.68% for the surgery department, and the ratio of repeating the test registration and test was higher in the internal medicine department.

5.4 Analysis Results of Payment Type

On average, the time consumed was 20 minutes for the general payment and 4.35 minutes for the unmanned payment. The following "Table 5" shows the time consumed on each activity.

Table 5. ***Processing time of 'general payment' and 'unmanned payment'***

Activity	Mean	Total
General Payment	20.0	61,582
Unmanned Payment	4.35	2,547

*unit: mins

Using the unmanned payment can save around fifteen minutes compared to using the general payment. If the hospital could recommend the patients to use unmanned payment, the waiting time of the general payment counter can be reduced while improving the customer satisfaction related to the payment.

VI. Conclusion

This research sought to comprehend the treatment process of the hospital using "R," the open-source analysis tool. The research conducted analysis related to gender, age, medical department, and payment method, and the result was as follows.

First, the age-specific comparison showed no difference in the number of activities related to the "tests," but the age group in the 20s to 30s showed a higher ratio of repeating the "test registration" and "tests" compared to the age group in the 60s to 70s. Additionally, the age groups showed differences in "general payment" and "unmanned payment."

Overall, the ratio of using the "general payment" was far higher, but the ratio of using the "unmanned payment" was higher as the age decreased. In the payment method-specific comparison, the average time consumed on the "unmanned payment" was 4.35 minutes, and the time consumed on the "general payment" was 20 minutes, indicating that people spend four times as much time when using the general payment. In other words, the use of the unmanned payment can save around fifteen minutes per patient, reducing the payment-related waiting time. Additionally, considering that the number of patients in the 60s to 70s is the highest, it is required that the hospital introduce more unmanned payment machines along with the guideline for senior patients to utilize the machines.

Second, the gender-specific analysis showed similar results in the treatment process, age group, and case number, but the throughput time showed a difference. The throughput time of the female group was shorter by six minutes compared to the male group; this may indicate that male patients tend to repeat the "test registration" and "tests" within one case, spending more time than female patients.

Despite the far bigger number of patients that visit the internal medicine department compared to the surgery department, the medical department-specific comparison did not show a significant difference in the frequency related to the "tests," and the ratio of visiting the surgery department decreased more drastically as the age increased.

Comparing the activities that appear on the process map, the "treatment" activity that the map of the internal medicine department include was included in the map of the surgery department. Additionally, the ratio of repeating the tests was higher in the internal medicine department (43.48%) compared to the surgery department (34.68%), and this indicates that the number of tests that a patient receives along the single visit is higher in the internal medicine department compared to the surgery department. Similarly, the analysis confirmed the difference in the treatment process between the medical department. For the smooth conduction of the medical process for patients, it is required that the hospital proposes a guideline for each medical department to the patients.

The data used in this research had very limited categories related to the "test registration" and "tests" activities (test registration, blood drawing, prescription execution). The varied test-related categories allow additional classifications of the activities and testing machine-specific analysis. Such an analysis can produce the index with which one can examine how efficient the machine is utilized by comparing utilization time, idle time and processing time.

ACKNOWLEDGMENT

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2018S1A3A2074932)

REFERENCES

[1] Kim, D. K., Kim, W. J., & Kim, M. R., Effects of Innovativeness on Customer Satisfaction in Long-Term Care Hospitals: The Effect of Internal Capacity and Location Strategy in Hospital. Journal of the Korea Academia-Industrial cooperation Society, Vol.20, No.3, pp.110-124, 2019.

[2] Kim, J. A., & Cho, E. Y., Computer use experience of nurses working in hospital. International Journal of Internet, Broadcasting and Communication, Vol.10, No.1, pp.31-39, 2018

[3] Cho, M. S., Song, M. S., & Yoo, S. Y., Analysis of Payment Using Process Mining. Proceeding of Korea Intelligent Information Systems Society, pp.129-132, 2013.

[4] Mans, R. S., Van der Aalst, W. M., & Vanwersch, R. J., Process mining in healthcare: evaluating and exploiting operational healthcare processes. Cham: Springer, pp.17-26, 2015.

[5] Kang, Y. S., & Lee, B. K., Process mining for business. Hannarae, 2016.

[6] Kim, K., & An, B., Effective Utilization of Data based on Analysis of Spatial Data Mining. The Journal of The Institute of Internet, Broadcasting and Communication, Vol.13, No.3, pp.157-163, 2013.

[7] Jung, J. D., Kim, S. H., & Yeo, H. J, Clinical Pathway Verification through Process Mining. Journal of the Korea Society of Computer and Information, Vol.23, No.4, pp.115-120, 2018.
[8] Van der Aalst, W. M., Reijers, H. A., Weijters, A. J., van Dongen, B. F., De Medeiros, A. A., Song, M., & Verbeek, H. M. W., Business process mining: An industrial application. Information Systems, Vol.32, No.5, pp.713-732, 2007.
[9] Lee, J. S., Kim, S. K., & Jung, J. Y., Analysis of Repair Service Process in a Telecom Company Using Process Mining. Proceeding of Korean Institute of Industrial Engineers, pp.595-602, 2014.
[10] Van der Aalst, W. M., van Dongen, B. F., Herbst, J., Maruster, L., Schimm, G., & Weijters, A. J., Workflow mining: A survey of issues and approaches. Data & knowledge engineering, Vol.47, No.2, pp.237-267, 2003.
[11] Choi, Y. H., Kim, Y. E., & Lee, Y. H., Prediction of Medical Service Process Using Data Mining and Process Mining. Proceeding of Korean Institute of Industrial Engineers, pp.1828-1840, 2015.
[12] Mans, R. S., Schonenberg, M. H., Song, M., van der Aalst, W. M., & Bakker, P. J., Application of process mining in healthcare–a case study in a dutch hospital. In International joint conference on biomedical engineering systems and technologies, Springer, Berlin, Heidelberg, pp.425-438, 2008.
[13] Lee, K. H., Lee, M. W., & Cha, J. S., Development of IoT-based data acquisition device (Gateway) for context-awareness of hospital facilities. The journal of the convergence on culture technology, Vol.3, No.4, pp.181-184, 2017.
[14] Jung, Y. G., Kim, H. W., & Kang, M. S., Predicting Discharge Rate of After-care patient using Hierarchy Analysis. International Journal of Advanced Culture Technology (IJACT), Vol.4, No.2, pp.38-42, 2016.
[15] Zhang, C., & Wan, L., The Extraction Process of Durative Persuasive System Design Characteristics for Healthcare-related Mobile Applications. International journal of advanced smart convergence, Vol.8, No.2, pp.18-29, 2019.