Mining Diagnostic Investigation Process

Sohail Imran[1] Dr. Tariq Mahmood[2]

Abstract

Diagnostic investigation process of healthcare is complex, medical practitioners' goal is to find methods of standardizing their diagnostic investigation processes to reduce the time and cost and optimize the quality of healthcare. The technique that can be applied to mine valuable and useful knowledge of diagnostic investigation process of their interest from stored data is process mining. Process does not consider dynamic and causal dependencies in processes. This characteristic of process mining can be effectively applied in diagnostic investigation. This technique becomes more helpful and valuable where some treatment failed to provide favouring evidence. We used process mining in this paper to mine efficient diagnostic investigation process flow for hepatitis patients. There are several advantages of using process mining approach which can boost the effectiveness diagnostic investigation processes.

Keywords: Process mining, healthcare, diagnostic investigation process, process flow.

1 Introduction

The emerging discipline evolved from the data mining is process mining [1]. The objective of process mining approach is the reengineering of a given business process model and visualized a certain aspect. Information systems are the backbone of executing process mining techniques.

The vital element of healthcare system is the correct diagnosis of a patient's condition, and the most important part of this process is diagnostic investigation. Diagnostic investigations provide a critical component of the patient's visit within the clinical management [2]. The information these tests provide affect the healthcare decisions. As the complexity of diagnostic investigation are increasing, conventional approaches to diagnose patient's condition are becoming increasingly inadequate [3].

Currently, there is a need of bridging gap in translational healthcare research between the diagnostic investigations associated with a particular disease and proving that patients who are tested for these diagnostic investigations have better outcomes than those who are not. Studies of diagnostic investigation accuracy are not sufficient to justify clinical use [4], [5]. Implementation of basic process mining techniques applied to real-life data and business problems is well documented [6]-[8]. Process mining can be effectively used to different business domain including healthcare. To improve diagnostic, treatment decisions and unaided human inference, process mining could be considered [9], [10].

Several laboratory units can be the part of healthcare diagnostic investigation process of patients. Laboratory test units are operated using specialized software modules, it is difficult to get data in uniform format for a specific disease.

[1]*PAF-Karachi Institute of Economics & Technology, Karachi, Pakistan | sohail@pafkiet.edu.pk*
[2]*PAF-Karachi Institute of Economics & Technology, Karachi, Pakistan | mahmood@pafkiet.edu.pk*

Event logs generated from these information systems are used to extract related knowledge [11]. There are several advantages of using process mining approach which can boost the effectiveness diagnostic investigation processes. In this paper, we have shown that it is possible to extract a useful and helpful picture of the real process without having in depth knowledge of the complex hospital process. We have presented different path maps. This can help not only medical practitioners but also forces health experts to unify diagnostic investigation process.

We have organized the remainder of this paper as follows: brief overview of process mining and the relationship between process mining and healthcare is presented in Section II. The description of the tool used to implement process mining in this paper is summarized in Section III. The process mining application is described in Section IV for healthcare diagnostic investigation using data, to obtain insights in an explorative manner. To get valuable insight the focus should not be on one aspect only. Therefore, we applied diagnostic investigations suggested to the patients on their visits and diagnostic investigation suggested by the medical practitioners. Finally, Section V concludes the paper and sketches a line for future research directions.

2 Related Work

A Process Mining

Process mining is the reverse engineering technique that uses event logs that are produced by information systems [7] [8]. Extraction of related knowledge from these recorded event logs is the basic idea of the application of process mining. Vast number of logs of events are generating against different activities of systems. These logs provided the basis for process mining. The aim of process mining is to extract valuable insights from the recorded event logs containing actual executed process instances and facts [12], [13].

Three process mining kinds are: i) Discovery, ii) Conformance analysis and iii) Extension. The first is no a priori model. It discovers previously unknown or undocumented processes from low-level events. It is performed when no model exists or when the quality of the existing documentation is poor. The event logs are analysed in order to discover the process. It generated documentation for the process. The Conformance analysis is an a priori model. It analyzes the deviation between the event log and the model. Its compares whether the current workflow conforms to the planned process. It generated discrepancies between the existing process and the model [7].

After locating discrepancies, processes are analyzed to suggest improvements in it. If these suggestions are beneficial to the process, the model is then reviewed to avoid these discrepancies using the information available at each node. Otherwise changes may then be made in the current process to conform to the model. The third kind of process mining is a prior model too. The goal of this kind of process mining is to find improvement in the existing model by extending the model. To find possible space of improvements in the existing model, the event logs are analyzed to get extension in the model or to get possible alternative paths in the workflow. Process models can be divided into two types: de jure process model and de facto process model. The first type of process model is normative, its purpose is to design future processes by incorporating enhancements in the existing model [8]. A second type process model is descriptive. Its purpose is to capture reality by mapping current processes to create a baseline for process improvements.

The event log data can also be divided into two types: post and pre mortem. The first type event log data is about completed cases which can be used for process enhancements. It does not change or affect the referred cases. The second type of event log data is about uncompleted cases which are currently in progress.

B Process Mining in Healthcare

Process mining can be used and implemented in Healthcare management for diagnosis and treatment [9]. To improve treatment decisions and unaided human inference, this technique can be used. The process mining application is very effective in diagnosing and detecting disease. It is useful in situation where some treatment failed to provide favouring evidence.

Optimal allocation of available healthcare resources is another key area where process mining can play an important role. Apart from getting benefit of process mining in healthcare, data mining is also effective in discovering process flow of treatment of diseases. Process mining can save cost and time involved in conventional techniques by analysing event log files to recognize treatment flow. Process mining provides view of interaction among different diagnostic investigation of a disease. This view of interaction detects anomalies before they become problems and discover the actual problems instead of immediately visible symptoms. It discovers medical practitioners' similar prescription process based on verifiable data [14].

Process mining establishes baselines for the existing diagnostic investigations and uses them to determine whether specific changes are effective or not. It helps in understand the entire process. Since process flow causes of the disease can be identified and seen in the context of the prescriptions, medical practitioners understand where and why change is needed.

The identification of diagnostic investigation patterns and the eventual satisfaction they result in can be used to improve overall patient satisfaction. In many cases prediction of diagnostic investigation can aid in designing proactive initiatives to reduce overall cost and increase patient satisfaction. To optimize the execution of processes, one of the best practices follow today is the standardization. Standardization of diagnostic investigation process could make the usage of health resources more effective that optimizes quality and efficiency of patient care [15], [16]. The application of process mining has several advantages to boost the effectiveness of medical diagnostic investigation processes [17].

This can help not only medical practitioners but also forces health experts to unify diagnostic investigation process. Along with the improvement in the quality of services using this approach, variations in daily diagnostic investigation practices can also be avoided. Effective resource management of healthcare is another benefit of this approach. This result in improved foresees and account the costs of treatment of patients.

3. Tool For Analysis

To analyze our data set, we used Fluxicon Disco process mining tool from the aspect of process diagnostics. The Disco is based on the Fuzzy miner. The Fuzzy Miner is a mining algorithm used to introduce the map metaphor to process mining. It includes seamless process simplification and highlighting of frequent activities and paths.

The data set is a collection of events applied for process mining is referred as event log. In data mining, an individual record represents a complete process instance but in process mining, an event is just an individual row. For the application of process mining, event log is the starting point. The analysis of data stored in event logs from the aspect of a process is the basic aim of the process mining. Process mining maps the data to a process view.

To apply process mining, the data need to fulfil certain minimum requirements. The three key elements are the identification of cases, activities and timestamps. The scope of the process is determined by the case and the activity determines the depth of level for the process steps. Every event that was executed in the process refers to a process instance or a case. Each case is a collection of multiple connected events. Process mining compares several executions of the process to one another. Another important requirement is timestamps. To get the right order of the events for each case, at least one timestamp is required otherwise the events are correctly ordered.

The most important analysis output in Disco is the process map. It generates the actual execution of the process. In the event log data, the time stamping and ordering of the stored activities is of most importance because Disco discovered the process path flows automatically on the bases of these two. The major advantage of using this approach is that we can extract auseful and helpful picture of the real process without having in depth knowledge of the complex process. Disco visualized the discovered process in a simple way: the stat of the process map is represented by a triangle symbol and a square symbol represents the end of the process. Boxes illustrated the activities and arrow shows the process flow between two activities. There are two types of arrows used in the tool. One solid line, discussed in earlier, and the dashed arrows visualized activities occurred at the very beginning or at the very end of the process. The numbers shown within activities and the arcs represent the frequency of processes and it also illustrated visually by the arrow thickness and the activity color. To generate the process map for our diagnostic investigation of patients, we used absolute frequency metric, based on total frequencies.

4 Process Mining For Healthcare Diagnostic Investigation

The application of process mining for healthcare diagnostic investigation will be shown in this section. In healthcare diagnostic investigation processes for the treatment of patients, several laboratory units can be involved and these laboratory test units are operated through specialized software modules, it is difficult to get data in uniform format for a specific disease. To get track of diagnostic investigation of patients for a specific disease, hospitals need to have integrated software modules of all laboratory test units.

Our event log contains information on the activities related to laboratory test prescribed to the patient on their visits by their medical practitioners in a hospital. The log contains events related to a patients diagnosed with hepatitis. The raw data used for this paper, sample data in Table1, contains information of 575 doctors, 10 diagnostic investigation and 1,313,177 records of hepatitis patients treated in 2010 to 2012. For hepatitis patients, the diagnostic investigation process is supported by several laboratory test units.

For this paper, event logs are extracted from the information systems database of the hospital. Each event in the data set is referred as a laboratory test of a patient. The rest of this section describes the application of process mining, for healthcare diagnostic investigation of hepatitis patients using data, to obtain insights in an explorative manner. So, the discovery part of process mining is our focal point not the conformance and extension part.

Table 1: A part of the event log

Visit ID	Doctor	Entry Date	Test Description
49359193	GJAV	31-Oct-10	Hemoglobin Haematocrit
49359193	GJAV	31-Oct-10	White Blood Cell Count
49359193	GJAV	31-Oct-10	S. Sodium
49359193	GJAV	31-Oct-10	S. Potassium
50617620	AJAF	3-Mar-10	Glucose Random
57921561	SMUN	11-Oct-10	Prothrombin Time

Figure 1 and Figure 2 show the process model of a patient visit-based view on the diagnostic investigation process and a doctor-based view on the diagnostic investigation process from the event log. To get valuable insight the focus should not be on one aspect only. Therefore,we have applied process mining from two perspectives: diagnostic investigation to patients on their visit and diagnostic investigation suggested by doctors.

In Figure 1 and Figure 2, diagnostic investigation process may consist of following activities: Neonatal TSH (NTSH), Complete Blood Count (CBC), Prothrombin Time (PT), APTT, Memoglobin Hematocrit (MH), Blood Urea Nitrogen (BUN), Creatinine (Cr), Serum Electrolytes (SE), Urine Detail Report (UDR), S. Sodium (SS), S. Potassium (SP), Hemoglobin Haematocrit (HH), Magnesium (Mg), Platelets (Pl), White Blood Cell Count (WBCC), SGPT, Glucose Random (GR). For a single instance of diagnostic investigation process, some of these activities might repeat or it is not necessary that all activities happen every time.

A On Patient Visit

In Figure 1, each case in the event log corresponds to a diagnostic investigation on patient visit. It can be seen that 30,974 different cases of the diagnostic investigation processes are there in the event log. There are two alternative process can be observed from the beginning. Since the beginning of the diagnostic investigation process, 2,868 cases ended after only one activity NTSH laboratory test. The other 28,106 cases perform activity CBC instead. It is clearly shown that 90.7% of the activities are started with the activity CBC.

Afterward, the process split into four alternatives paths. Out of four activities, diagnostic investigation process terminated after the prescription of HH laboratory test without proceeding to next activity. Another pattern can be noticed for the other alternative that loop backed to CBC laboratory test activity after performing PT and APPT laboratory test activities. The third alternative, started from the activity BUN laboratory test, is the central path flow in this

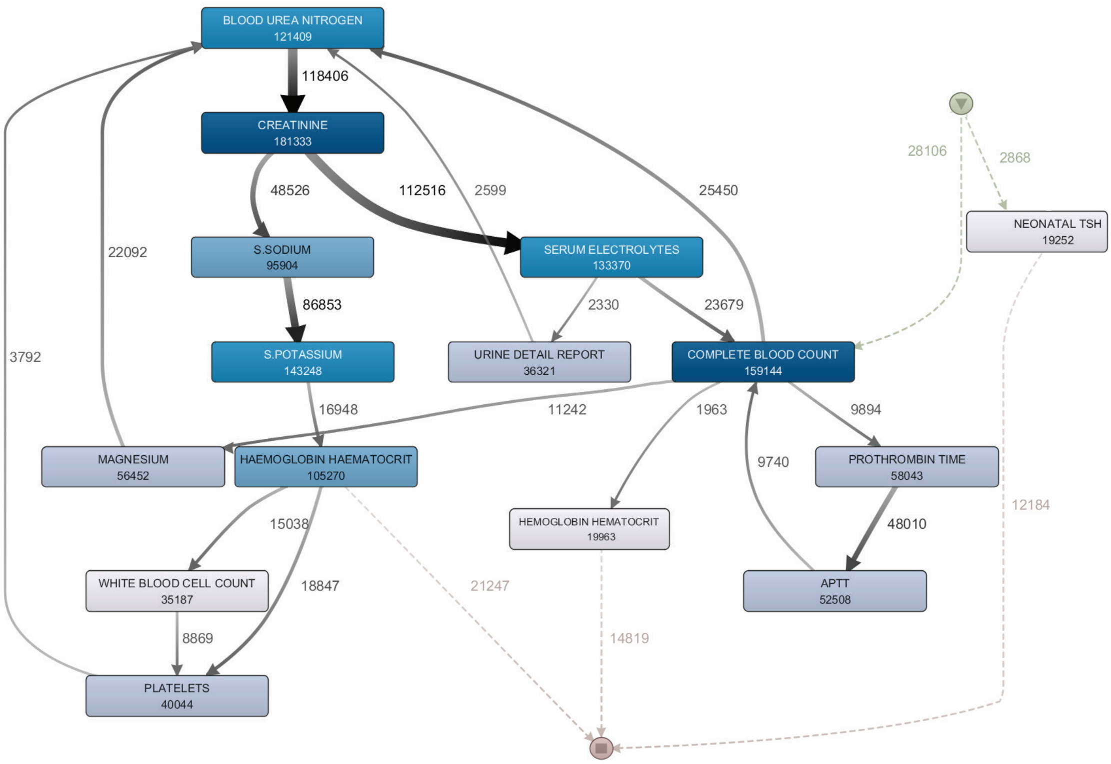

Figure 1: A patient visit-based view on the diagnostic investigation process

segment of diagnostic investigation process that can be visualized by the thickest arrow with a weight of 118,406 cases between BUN and Cr.

Overall, activity Cr laboratory test is most executed activity (in total 181,333 times). This activity further splits into two dominant loops. First, Cr-SE-CBC[UDR]-BUN-Cr loop and second Cr-SS-SP-HH-Pl[WBCC]-BUN-Cr loop. The fourth alternative is actually same process as third alternative with an additional activity Mg laboratory test, before joining it to the third alternative. 21,247 case of the second loop of third alternative completed and end with activity HH laboratory test.

The process map of Figure 1 discovered important results to diagnose hepatitis in the form of two important flows. The first flow CBC-BUN-Cr-SE-CBC[UDR]-BUN and the second CBC-BUN-Cr-SS-SP-HH-Pl[WBCC]-BUN identified most frequent laboratory test activities to diagnose the disease. Another analysis result was the detection of termination decision of disease treatment. The process terminated after NTSH and HH that showed activities helpful in decision making.

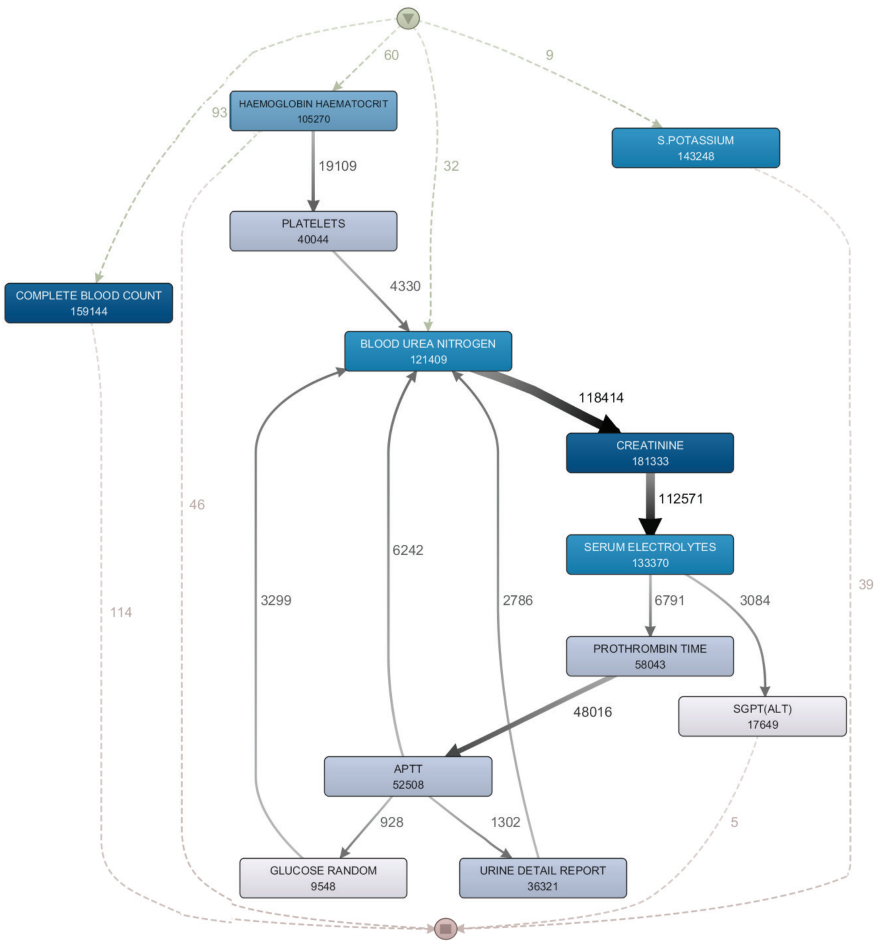

Figure 2: A doctor-based view on the diagnostic investigation process

B By Doctor

In Figure 2, each case in the event log corresponds to a diagnostic investigation by doctor. There are four alternative diagnostic investigation process can be observed from the beginning. Out of four, two ended after only one activity. First one ended after SP laboratory test, second after CBC.

The third alternative, started from HH laboratory test activity, further split into two fragments; first fragment ended in the similar fashion as of first and second alternatives. The other fragment merged with the fourth alternative path that started with BUN laboratory test, after performing an additional activity Pl laboratory test. Afterward, the process continued Cr laboratory test activity. The second fragment of the third alternative, started from the activity HH laboratory test and merged with the activity BUN laboratory test after completing Pl laboratory test activity, is the central path flow in this segment of the diagnostic investigation process.

A loop backed pattern to BUN laboratory test activity can be noticed after merging fourth alternative with the second fragment of third alternative. It is clearly shown that 99% of the activities are started with the activity HH. Over all, activity Cr laboratory test is most executed activity (in total 118,414 times). This activity proceeds to SE laboratory test and then further splits into two. First one ended after performing SGPT laboratory test activity and the second one followed to APTT laboratory test activity. This activity further splits into two loops. First, APTT-UDR-BUN loop and second APTT-GR-BUN loop.

The process map of Figure 2 discovered important results to diagnose hepatitis in the form of an important flow. The flow BUN-Cr-SE-PT-APTT-[UDR|GR]-BUN is identified as most frequent test laboratory activities to diagnose the disease. Another analysis result was the detection of termination decision of disease treatment. The process terminated after SP, CBC, SG and HH that showed activities helpful in decision making.

5 Conclusion

This research paper showed that the effective application of process mining approach to the diagnostic investigation of healthcare domain data. We applied process mining from two perspectives to extract relevant knowledge: diagnostic investigation to patients on their visit and diagnostic investigation suggested prescribed by doctors. For these two aspects, we presented some initial results. We have shown that without having any existing diagnostic investigation process model and in depth knowledge of the complex daily diagnostic investigation processes, we can extract the objective picture of the real diagnostic investigation process using process mining approach.

In addition, to get valuable insight the focus should not be on one aspect only. Therefore, we applied diagnostic investigations suggested to the patients on their visits and diagnostic investigation suggested by the medical practitioners. We have presented different path maps. This can help not only medical practitioners but also forces health experts to unify diagnostic investigation process. Along with the improvement in the quality of services using this approach, variations in daily diagnostic investigation practices can also be avoided.

The process map discovered important results to diagnose hepatitis in the form of important flows. We identified most frequent test laboratory activities to diagnose the disease. Another analysis result was the detection of termination decision of disease treatment which is helpful in decision making. This paper recommended that a medical practitioners' group be established to discuss how best to develop a diagnostic investigation mechanism for the evaluation of new laboratory investigations.

This research has presented diagnostic investigation process for hepatitis patient. At the same way, these results evidence some facilities provided to the expert to guide their process to the knowledge about other diseases. Also we have data of one hospital only; using data of multiple hospitals can be used to compare diagnostic investigations of multiple hospitals. In future the application of process mining will be very effective to present generalized diagnostic investigation process using big data and NoSQL technologies.

References

[1] W. M. P. van der Aalst, “Process Discovery: Capturing the Invisible,” IEEE Computational Intelligence Magazine, vol. 5, no. 1, 2010, pp. 28–41.

[2] P.J.F. Ramírez, R.T. Rodríguez, F.D. Olivera, V.M. Morejón, “Component to decision making in health. A social network analisis approach from process mining,” Revista Cubana de Informática Médica, 8 (1), 2016, pp. 46-63.

[3] C. Kulatunga-Moruzi, "Investigating the diagnostic process: The coordination of diagnostic rules and clinical experience," ETD Collection for McMaster University, 2007.

[4] G. G. Raab, L.J. Logue, “Medicare Coverage of New Clinical Diagnostic Laboratory Tests: The Need for Coding and Payment Reforms,” Clinical Leadership & Management Review, vol. 15, no. 6, 2001.

[5] N. Foshay, C. Kuziemsky, “Towards an implementation framework for business intelligence in healthcare,” International Journal of Information Management, 2014.

[6] M. Dumas, W.M.P. van der Aalst, A.H. Hofstede, "Process Aware Information Systems: Bridging People and Software Through Process Technology," Journal of the American Society for Information Science and Technology, vol. 58, no. 3, 2007, pp. 455–456.

[7] W. M. P. van der Aalst, A. Andriansyah, A.K. Alves de Medeiros, F. Arcieri, T. Baier, T. Blickle, J.C. Bose, P. Van den Brand, R. Brandtjen, J. Buijs, "Process Mining Manifesto," in Business Process Management 2011 Workshops Proceedings, 2012, pp. 169–194.

[8] R.S. Mans, M.H. Schonenberg, M. Song, W.M.P. van der Aalst, P.J.M. Bakker, "Application of Process Mining in Healthcare – A Case Study in a Dutch Hospital," Biomedical Engineering Systems and Technologies, 2008, pp. 425–438.

[9] H.C. Koh and G. Tan, “Data Mining applications in healthcare”, Journal of Healthcare Information Management.,vol. 19, no.2, 2005, pp. 64-72.

[10] W.M.P. van der Aalst, B.F. van Dongen, J. Herbst, L. Maruster, G. Schimm, A.J.M.M. Weijters, “Workflow Mining: A Survey of Issues and Approaches,” Data and Knowledge Engineering, vol. 47, no.2, 2003, pp. 237-267.

[11] A.J.M.M. Weijters, W.M.P. van der Aalst, "Process Mining: Discovering Workflow Models from Event-Based Data," 13th Belgium-Netherlands Conference on Artificial Intelligence, 2001, pp. 283–290.

[12] W.M.P. van der Aalst, V. Rubin, B.F. Dongen, E. Kindler, C.W. Günther, "Process Mining: A Two-Step Approach to Balance Between Underfitting and Overfitting," Software and Systems Modeling, vol. 9, no. 1, 2010, pp. 87–111.

[13] A. Rozinat, W.M.P. van der Aalst, "Conformance Checking of Processes Based on Monitoring Real Behavior," Information Systems, vol. 33, no. 1, 2008, pp. 64–95.

[14] J. Tanawat, A. Poohridate, K. Kwanchai, I. Sarayut, P. Wichian, "Conformance analysis of outpatient data using process mining technique," ICT and Knowledge Engineering, 15th International Conference Proceedings, 2017.

[15] N.R. Every, J. Hochman, R. Becker, S. Kopecky, C. P. Cannon, "Critical pathways. a review," National Library of Medicine and The National Institutes of Health, 2000, pp. 461–465.

[16] M. Erraguntla, P. Tomasulo, K. Land, H. Kamel, M. Bravo, B. Whitaker, R. Mayer, S. Khaire, "Data Mining to Improve Safety of Blood Donation Process," System Sciences (HICSS), 2014 47th Hawaii International Conference, 2014, pp. 789-795.

[17] H. Laihonen, "Knowledge structures of a health ecosystem," Journal of Health Organization and Management, 2012, pp. 542–558.