ADDIS ABABA UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCE

SCHOOL OF INFORMATION SCIENCE

IMPROVING HEALTHCARE SYSTEM THOUGH PROCESS MINING AT TIKUR ANBESSA SPECIALIZED HOSPITAL

BY

MESELE AWULACHEW (ID.No: GSR/0335/15)

OCTOBER, 2024

ADDIS ABABA, ETHIOPIA

ADDIS ABABA UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCE

SCHOOL OF INFORMATION SCIENCE

IMPROVING HEALTHCARE SYSTEM THOUGH PROCESS MINING AT TIKUR ANBESSA SPECIALIZED HOSPITAL

A Thesis Submitted to School of Graduate Studies of Addis Ababa University in Partial Fulfillment of the Requirements for the Degree of Master of Science in Information System

By: MESELE AWULACHEW

Advisor: MILLION MESHESHA (PhD)

OCTOBER, 2024

ADDIS ABABA, ETHIOPIA

ADDIS ABABA UNIVERSITY
COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCE
SCHOOL OF INFORMATION SCIENCE

IMPROVING HEALTHCARE SYSTEM THOUGH PROCESS MINING AT TIKUR ANBESSA SPECIALIZED HOSPITAL

By: MESELE AWULACHEW

Name and Signature of Members of the Examining Board

Million Meshesha (PhD)	____________________	____________________
Advisor	Signature	Date
____________________	____________________	____________________
Examiner	Signature	Date
____________________	____________________	____________________
Examiner	Signature	Date

DECLARATION

This thesis has not previously been accepted for any degree and is not being concurrently submitted in candidature for any degree in any university.

I declare that the thesis is a result of my own investigation, except where otherwise stated. I have undertaken the study independently with the guidance and support of my research advisor. Other sources are acknowledged by citations giving explicit references. A list of references is appended.

Signature: _______________________________________

Mesele Awulachew

This thesis has been submitted for examination with my approval as university advisor.

Advisor Signature: __________________________________

Million Meshesha (PhD)

DEDICATION

This thesis is dedicated to the memory of my beloved father, Awulachew Zimbele

ACKNOWLEDGEMENT

First and foremost, I want to express my sincere gratitude to God for allowing me to complete this phase of my research. His guidance and blessings have been pivotal in this journey.

I would like to extend my deepest gratitude to my advisor, Dr. Million Meshesha. His unwavering support, insightful advice, and constructive criticism have been invaluable in guiding this research. Dr. Million's expertise and dedication have significantly contributed to the improvement and success of this study. His mentorship has been a source of inspiration, and I am profoundly grateful for his guidance throughout this research journey.

I am also deeply grateful to the staff of the Internal Medicine Department and the ICT Office at Tikur Anbesa Specialized Hospital. Their kindness in granting me access to the necessary research data and other information has been instrumental in the completion of this project.

My gratitude extends to the Department of Management, College of Business and Economics, Addis Ababa University, for providing me with the opportunity to attend the postgraduate program and for their sponsorship.

Finally, I acknowledge the support received from Addis Ababa University, particularly from the College of Natural and Computational Sciences, Department of Information Science. Their facilitation and endorsement of this research project have been greatly appreciated.

ABSTRACT

The quality of hospital services depends on the effective execution of healthcare processes, which encompass a variety of clinical and non-clinical activities performed by diverse resources. These processes are dynamic, complex, and multi-disciplinary, necessitating a deep understanding of improvement. Process mining offers promising techniques for visualizing and analyzing healthcare processes to enhance efficiency and quality. This study therefore aims to apply process mining techniques to discover healthcare process models, identify deviations and inefficiencies, and optimize resource allocation within the hematology department of a healthcare organization. The process model followed in this research includes data extraction from Tikur Anbessa Specialized Hospital (TASH), data preprocessing using aggregation, temporal approach and simple heuristic, process discovery using heuristics mining and inductive mining, and model evaluation based on fitness, precision, generalization, and simplicity. Control flow, performance and organizational analyses are also conducted, followed by validation of findings through expert collaboration.

The analysis highlights the most common pathway for hematology patients begins with a laboratory request, followed by a laboratory test, then a hematology diagnosis, and finally a prescription, highlighting the interconnectedness of these processes. However, discrepancies between the number of laboratory requests and completed tests, coupled with an average test duration of 32 days, significantly above World Health Organization (WHO) benchmarks, reveal inefficiencies, particularly in resource allocation. Comparative analysis using heuristic and inductive miners demonstrated the inductive miner showing superior fitness, precision and simplicity, with the heuristic miner achieves a slightly higher in generalization. Social Network Analysis (SNA) identified strong interdepartmental interactions, especially in the diagnosis and radiology departments. The proposed process improvement framework was well-received, achieving an overall mean evaluation score of 4.2 and a Cronbach's alpha of 0.747, indicating its reliability. These findings emphasize the complexity of healthcare processes and the importance of continuous improvement through integrated systems. Future research should address challenges in data quality issue to further enhance the utility of process mining in healthcare settings.

Keywords: Process Discovery, Process Mining, Process Discovery Techniques, Healthcare Service, Process Improvement

TABLE OF CONTENTS

LIST OF ACRONYMS

BPMN	Business Process Modelling and Notation
CPOE	Computerized Physician Order Entry
CSV	Comma-Separated Value
ED	Emergency Department
EHR	Electronic Health Record
EMR	Electronic Medical Records
ERP	Enterprise Resource Planning
HIS	Hospital Information System
IEEE	Institute of Electrical and Electronics Engineers
KPIs	Key Performance Indicators
MRN	Medical Record Number
PACS	Picture Archiving and Communication Systems
PM2	Process Mining Project Methodology
PMS	Practice Management Systems
ProM	Process Mining Framework
SNA	Social Network Analysis
TASH	Tikur Anbessa Specialized Hospital
WHO	World Health Organization
XES	eXtensible Event Stream
YAWL	Yet Another Workflow Language

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

1.1. Background of the Study

Healthcare processes must be carried out properly and efficiently to provide high-quality hospital services. A healthcare process can be defined as the series of activities that healthcare providers and organizations perform to deliver healthcare services to patients [1]. These processes are supported by clinical and non-clinical activities, executed by different types of specialists and professionals (physicians, nurses, technical specialists, dentists, clerks), and can vary from one organization to another [2]. Although healthcare is typically associated with hospitals, there are many care processes in other types of organizations, where care is provided at home, rehabilitation centers, and nursing homes. It is indicated that the healthcare process is not always linear and is often highly dynamic, complex, ad-hoc, and increasingly multi-disciplinary [3]. Patients may move back and forth between stages as their condition changes [4]. For example, a patient who is being treated for cancer may also need to receive supportive care to manage the side effects of treatment, making them difficult to manage and improve. Understanding the characteristics of healthcare processes is essential for improving the quality and efficiency of care delivery. By identifying the factors that contribute to variability, interdependency and high stakes, healthcare organizations can develop strategies to mitigate these challenges [5].

In light of the increasing use of information systems in business process work, process mining has emerged as a promising technology to visualize, analyze, and improve business processes [6]. Process mining focuses on analyzing end-to-end processes. It becomes a research issue because of the growing availability of event data and new process discovery and conformance-checking techniques [7]. This technology has become available only recently but is mature enough to be applied to care processes of any type and of any complexity [2]. Process models were formerly usually created by hand without the use of event data. But actions taken by humans, computers, and software all leave traces in event logs. Event logs can be used to conduct three types of process mining [8]. The first one is a ***discovery*** technique that takes an event log and produces a model using techniques like the alpha miner, heuristic miner, inductive miner, fuzzy miner, and genetic process mining, focusing on control flow to understand the sequence of activities. The other is ***Conformance*** process mining by comparing an existing process model with an event log of the same process, identifying deviations through methods such as replaying log events on the model and ensuring compliance. Lastly, process ***enhancement*** involves leveraging data from event logs containing information about the actual process to expand or enhance an already-existing process model, utilizing techniques like dotted chart analysis and decision mining for performance and organizational perspectives.

Healthcare organizations face numerous challenges in delivering high-quality patient care while controlling costs and improving operational efficiency [9]. One of the key challenges is in understanding and optimizing the complex processes and workflows involved in patient care, which can be difficult to observe and analyze due to the dynamic and decentralized nature of healthcare delivery. To be able to understand whether a healthcare organization achieves its goals of providing timely, cost-effective, and quality medical services, we need to analyze its healthcare processes [10].

Process mining, a specific application of data mining, offers a powerful tool for healthcare organizations to address this challenge by providing a systematic and data-driven approach to analyze and optimize their processes. Process mining can help healthcare organizations understand their processes better, identify bottlenecks and inefficiencies, and develop improvement plans [11]. The application of process mining is far more interesting, as it is not limited to re-discovering what we already know, but it can be used to unveil previously hidden knowledge [12]. Despite the growing recognition of the importance of business process analysis and process mining, especially within the healthcare sector, this technology remains largely underutilized, particularly in regions such as Africa and Latin America [13] [1]. The lack of adoption of this technology poses significant challenges for health providers, hindering their ability to leverage process mining insights effectively. Consequently, the potential benefits of evidence-based process improvement in healthcare will remain unrealized until these issues are addressed.

Applying process mining in healthcare presents several challenges [8]. Healthcare process models are complex due to diverse patient needs and varying granularity of data [55]. Incomplete event logs, distributed data sources, and concept drift add to the complexity [81]. Involving domain experts during the abstraction stage and ensuring data quality are resource-intensive but crucial for accurate analysis. Visualization of complex models and effective representation of variable healthcare processes are also challenging [55]. Additionally, privacy and security concerns must be addressed to protect patient information [11]. Despite these challenges, process mining offers significant potential for improving healthcare services by providing insights into underlying processes, identifying bottlenecks, as well as enhancing overall efficiency and patient satisfaction [12].

1.2. Statement of the Problem

In Ethiopia, since the hospitals are limited in number and scarce supply of equipment, it makes the treatment process more complex and inefficient. In particular, Tikur Anbessa Specialized Hospital (TASH) is the largest and the main governmental Hospital. Hence patients are referred to this hospital from all over the country. As a result, the hospital is always crowded with patients, making the provision of healthcare services very challenging [14] [15]. These challenges can be categorized into various aspects. One of the primary challenges faced by TASH is inadequate infrastructure [15] [16]. Insufficient infrastructure leads to overcrowding and compromises the quality of care provided to patients. Limited resources also hinder the hospital's ability to handle emergencies and provide timely treatment. Another significant challenge is the shortage of qualified healthcare professionals [15] [16]. This shortage can result in increased workloads for existing staff, leading to exhaustion and decreased efficiency. Insufficient staffing levels also impact patient care as it becomes challenging to provide personalized attention and timely treatment. Several patient-related factors contribute to the difficulty in providing healthcare services at TASH. These include low health literacy levels among patients, delayed presentation for treatment, and non-compliance with prescribed medications or follow-up appointments [17] [18].

To fix these constraints, it is necessary to identify the problem and improve the healthcare process by applying process mining techniques. Process mining has been applied in various studies in healthcare institutions and hospitals for discovering process models from event logs [13] [19] [20], for conformance checking [20] [21], and for enhancing efficiency of workflows [22] [23]. Among the various perspectives case studies were applied for investigating control flow perspective [23] [25], performance perspective [20] [21] and organizational perspective [22] [23]. The result of the studies shows that process mining offers substantial benefits to the healthcare system by enabling process optimization, improving patient care, ensuring compliance with regulations, and supporting evidence-based decision-making.

The previous case studies focused on applying process mining techniques to structured healthcare processes. The researchers recommended the application of these techniques in complex and unstructured healthcare processes for a more comprehensive understanding of bottlenecks within the workflow, as well as to improve outcomes. It is important to acknowledge that our understanding and management of health are influenced by various factors such as culture, society, economics, environment, and individual differences. Each of these elements plays a critical role in how health is perceived, prioritized, and addressed within different communities. For instance, cultural attitudes can affect health behaviors and practices, while socioeconomic status may determine access to healthcare resources. Environmental factors, such as living conditions and exposure to pollutants, further influence health outcomes. These contextual factors significantly shape health outcomes, healthcare practices, and public health policies [26] [27] [28]. In another research done by Rogers, De Brún, and McAuliffe [29], various factors within the context, such as organizational culture, leadership styles, available resources, patient demographics, geographical location, and social norms, can influence how healthcare processes unfold. For example, a supportive organizational culture and strong leadership can foster collaboration and innovation, leading to improved patient outcomes. Conversely, limited resources or unfavorable social norms may create barriers to accessing care and hinder the implementation of best practices. By examining these contextual influences, we can gain insights into the complexities of healthcare systems and develop strategies to enhance their effectiveness, ultimately improving the quality of care provided to patients.

Healthcare systems worldwide, particularly in resource-limited settings like Tikur Anbessa Specialized Hospital (TASH), face challenges related to inefficiencies, bottlenecks, and process deviations that compromise service quality and patient satisfaction. The complexities of healthcare processes, including patient flow, treatment timelines, and the coordination of medical activities, make it difficult to ensure that care is delivered in a timely, efficient, and effective manner [2]. Despite advancements in medical technologies, healthcare organizations often struggle with optimizing their internal processes due to limited visibility into how these processes unfold in practice [10].

Process mining, a data-driven technique that enables the extraction of insights from event logs, offers the potential to address these inefficiencies by uncovering the actual flow of healthcare activities. This includes identifying critical pathways, deviations, and performance bottlenecks that disrupt the smooth delivery of care. However, a significant gap remains in applying process mining methodologies specifically to healthcare processes, where unstructured and complex workflows add layers of difficulty [11].

Thus, the problem lies in the lack of a comprehensive, data-driven framework that can both extract and leverage insights from healthcare processes to inform strategic improvements. This necessitates a structured approach for selecting the most suitable process mining techniques, identifying critical inefficiencies, and proposing an actionable framework to improve healthcare services [24]. Addressing these issues would not only enhance the efficiency of care delivery but also contribute to improved patient outcomes and satisfaction. It is therefore the aim of this study to apply process mining to investigate the hematology care pathways and improve the efficiency and quality of patient care in the TASH.

To investigate and solve the above-mentioned process and health care pathways related problems, the current study tries to address the following research questions,

- What are the suitable process mining techniques for extracting insight in the healthcare process?
- What are the critical pathways, deviations, and bottlenecks identified in the healthcare business process?
- What is the best framework proposed to improve healthcare service by using extracted insight in the healthcare process?

1.3. Objective of the Study

1.3.1. General Objective

The general objective of the study is to analyze healthcare system through process mining techniques and propose a framework to improve healthcare service at TASH.

1.3.2. Specific Objectives

In order to achieve the general objective of the study, the following specific objectives are formulated.

- To identify suitable process mining methods and techniques
- To discover care flow models from the event log using model discovery techniques

- To discover the relationship between departments using social network mining techniques
- To identify the deviation, inefficiency and bottlenecks in healthcare processes at TASH
- To develop and evaluate healthcare service improvement framework

1.4. Significance of the Study

This study is crucial for TASH as it highlights inefficiencies and bottlenecks in the hematology treatment process, particularly the prolonged sojourn time. By adopting the proposed process improvement framework, TASH can improve the flow of patients, reduce waiting times, and better utilize its resources, ultimately leading to more efficient and effective healthcare provision.

Hematology specialists at TASH will benefit from a clearer understanding of the treatment process and its variations, as revealed by the process model. The analysis of control flow, performance, and organizational aspects will enable them to identify and address inefficiencies in their workflows. This will also enhance collaboration among departments and roles within the hospital, leading to more coordinated and effective patient care.

The study's outcomes directly contribute to improving patient satisfaction by streamlining the hematology treatment process. By minimizing delays, and optimizing the flow of patients through the treatment pathway, the hospital can better meet patient expectations. Faster and more efficient care will likely lead to improved health outcomes and a more positive experience for patients, reinforcing their trust in the hospital's services.

For researchers in the field of healthcare process optimization, this study offers a valuable case study in the application of process mining in a real-world hospital environment, particularly in resource-constrained settings like TASH. The methodological approach provides a robust framework that can be replicated or adapted for similar studies in other healthcare environments. The research also contributes to the growing body of knowledge on the application of process mining in healthcare. Furthermore, the study's findings and the proposed framework offer a basis for future research on process improvement.

1.5. Scope and Limitation of the Study

This study is conducted at TASH and focuses exclusively on the hematology treatment process specifically for outpatients. It is not covered inpatient treatment due to the limitations of the current healthcare information system, which only supports outpatient services and cannot provide inpatient data. Our study covers data from 2020 to 2024. This timeframe is crucial as the IWKET was launched in 2020, whereas Medweb was implemented earlier. By extracting and integrating data from both systems, we are able to analyze the complete patient journey throughout this period. The study uses process mining techniques to evaluate the treatment process from three main perspectives such as control flow, performance, and organizational. It does not focus on individual cases but rather aims to give a broad overview of the overall journey of hematology patients.

A significant limitation of this research is the lack of direct access to event logs from the hospital information systems, as the system is not process-aware. This necessitated the creation of an event log through various preprocessing steps, which could introduce data quality issues. Another limitation was that the healthcare information system at TASH is implemented only for outpatients. Consequently, the study is unable to analyze critical inpatient processes in hematology treatment, which may have impacted the comprehensiveness of the findings.

1.6. Organization of the Study

The first chapter provides a comprehensive overview of the study, encompassing the background of healthcare processes, the concept of process mining, and its application in healthcare. It also addresses the problem statement, outlines the general and specific objectives of the study, defines the scope and limitation of the research, and discusses its significance.

Chapter two presents a literature review focused on healthcare service processes and an overview of process mining. It explores the different types of process mining, its applications, the challenges in its application, and related works. It also reviews various process mining techniques and tools.

Chapter three discusses methodology and explores specific process mining methods and techniques selected for this study, including heuristic mining, inductive mining, and social network analysis methods.

Chapter four primarily addresses the stages of data extraction and preprocessing. It covers the integration of data from different sources into a single data during the extraction process and the reduction and filtering out of irrelevant data during the preprocessing phase.

Chapter five focuses on the experimental phase of the study. It discusses the stages of process mining experimentation, such as process model discovery, analyzing the models from different perspectives, evaluating the quality of process models, proposing healthcare improvement framework and evaluating the results.

Chapter six offers the conclusion and recommendations. It concludes the findings of the study, the research contribution, suggests areas for future research, and provides recommendations for organizational improvements based on the research findings.

CHAPTER TWO

LITERATURE REVIEW

In this chapter, review of literature is presented to provide highlight of the process involved in health care, followed by an overview and types of process mining. Finally, related works are summarized to show the gap in the current studies.

2.1. Overview of the Healthcare

In order to improve an individual's quality of life, healthcare can be defined as the diagnosis, treatment, and prevention of diseases. Hospitals are commonly linked with healthcare. However, various care processes can also be found in other kinds of organizations. These processes may involve a variety of professions [2].

2.1.1. Healthcare Processes

A number of activities are involved in healthcare process with the goal of enhancing patients' health through the diagnosis, treatment, and prevention of all diseases. These processes can differ between organizations and comprise several clinical and nonclinical tasks carried out by a range of professionals, including doctors, dentists, nurses, technicians, and clerks [43]. While hospitals are commonly linked with healthcare, a variety of care processes also take place in alternative settings such as homes, rehabilitation centers, and nursing facilities [2].

In the healthcare industry, there are two primary categories of processes such as organizational processes and medical treatment procedures [43]. Clinical treatment procedures, sometimes referred to as medical treatment processes, are closely related to the patient and follow a diagnostic-therapeutic cycle that includes observation, reasoning, and action. To handle case-specific decisions that are made by analyzing patient-specific information, the diagnostic-therapeutic cycle significantly depends on medical expertise. Administrative or organizational processes, on the other hand, are general process patterns that assist medical care processes at organizational level. They are intended to coordinate medical therapy across several individuals and organizational units rather than being customized for a specific medical condition. Exam requests and patient scheduling are two examples of organizational procedures. These processes are not all simple. They are executed in the continually changing environment when compared to others [3]. According to Ginter, Duncan, and Swayne [44], the healthcare system has experienced significant transformation and is expected to face even more substantial changes in the future. Their study indicates that healthcare organizations will need to adapt to various types and levels of change, including those related to legislation/politics, economics, social/demographics, technology, and competition.

2.1.2. Characteristics of Healthcare Processes

According to Rebuge and Ferreira [3], healthcare processes and their related activities have specific features related to their dynamic nature, complexity, and multi-disciplinary context. Healthcare practices frequently show considerable variety and heavily depend on complicated judgments made by skilled experts, like doctors, who often operate with a high level of independence [11]. In general, healthcare processes are recognized to have the following characteristics:

- **Healthcare processes are highly dynamic:** Healthcare procedures are highly variable due to simultaneous activities [11], evolving protocols, new technologies, and emerging diseases. Patient preferences and characteristics further influence treatment decisions and outcomes [3]. Patients also have the authority to accept or decline treatments based on personal beliefs and perceptions [11].
- **Healthcare processes are highly complex**: Healthcare operations are highly complex due to the involvement of multiple stakeholders, diverse medical fields, and rapidly advancing technologies. The complexity is further increased by vast amounts of diverse data, such as electronic health records and genomic information [45]. Medical decision-making is intricate, involving patient-specific data, medical expertise, and varying patient responses to treatments [3].
- **Healthcare processes are increasingly multi-disciplinary**: Healthcare institutions now feature more specialized departments and operate within interconnected networks, requiring collaboration across diverse professionals with varying expertise [3]. This collaboration is essential for patient-centered care, positively influencing care quality, job engagement, and reducing errors [46]. Effective teamwork also lowers healthcare professionals' intention to leave their positions [47].
- **Healthcare processes are ad hoc**: Healthcare processes are highly variable and unpredictable due to the specialized knowledge and independent judgment of professionals , along with rapid

advancements in medical technology and treatments [3]. This dynamic environment makes it challenging to maintain standardized workflows [48].

- **Healthcare processes value the infrequent behavior**: Infrequent behaviors in healthcare can reveal inefficiencies and challenges, making it crucial for professionals to recognize and manage them for comprehensive patient care [48]. Ignoring these behaviors risks misdiagnosis, delayed treatment, and inadequate care [49].

2.1.3. Clinical Pathways

Terms like clinical pathways, care flows, and healthcare processes are often used interchangeably [41]. A clinical pathway is a collaborative tool that standardizes interventions and treatment sequences to improve patient outcomes and resource utilization [50]. It aims to ensure consistent, high-quality care, predictable hospital stays, and better coordination across disciplines [51]. A clinical pathway can be supported by documenting activities in medical records and using clinical information systems for analysis and guidance. Refinement and standardization enable the creation of new pathways linked to existing ones [52].

According to Napolitano [53], clinical pathways typically include the following key components:

- **Timeline**: provides a precise schedule for interventions and activities, ensuring efficient patient care by setting specific timeframes for tests, treatments, and other interventions.
- **Categories of Care or Activities and Their Interventions**: Patient care activities including initial assessment, diagnostic testing, medication treatments, surgical procedures, rehabilitation, and discharge planning.
- **Intermediate and Long-Term Outcome Criteria**: It outlines the anticipated outcomes and significant checkpoints during various phases of the patient's treatment process, supporting systematic tracking of patient advancement.
- **Variance Record:** The variance documentation element is essential for recording deviations from the standard care pathway, allowing healthcare professionals to analysis to identify factors leading to divergences and improve the quality of care.

2.1.4. Different Levels of Healthcare

According to Mans, Van der Aalst, and Vanwersch [2], regarding the organization of healthcare, there is a distinction between three levels of care. Each level corresponds to a particular patient's needs. Primary care is the first point of contact for health services and focuses on preventive care and the treatment of common ailments through routine check-ups. Secondary care requires referrals from primary providers and involves medical specialists addressing more complex conditions. Tertiary care is hospital-based care and deals with severe or rare conditions through complex procedures, demanding specialized equipment, and a team of experts.

2.1.5. Classification of Healthcare Processes

Healthcare process execution can vary depending on the complexity of patients. While similar process execution may be observed in homogeneous patient groups, diverse outcomes in processes can occur with complex patients. Understanding the various characteristics of healthcare processes is important for efficient management and delivery of care [2]. Figure 2.1 illustrates the primary types of healthcare processes to offer insight into their varied nature and impact on patient care.

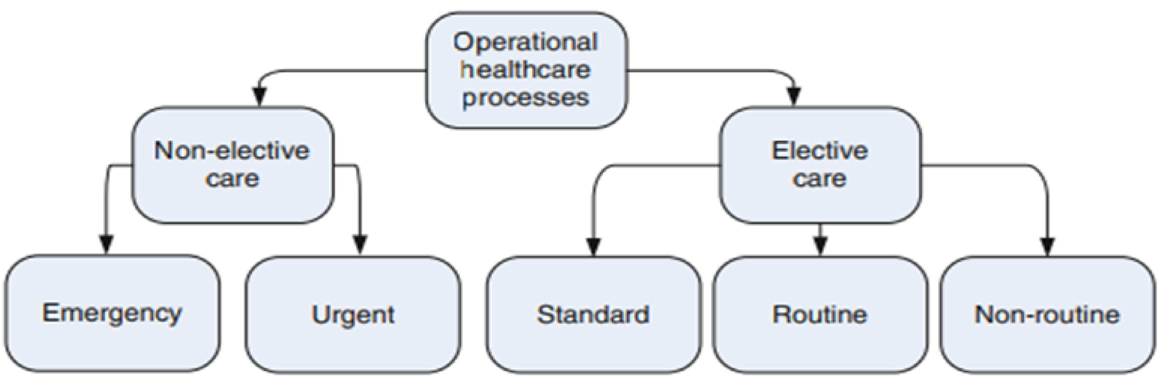


Figure 2.1: Classification of organizational healthcare processes [2]

The organizational healthcare process classification focuses on planning and managing operational processes like scheduling and accommodations, rather than individual medical decisions. As shown in Figure 2.1, it distinguishes between non-elective care, for urgent treatments, and elective care, which can be scheduled in advance. According to Lillrank and Liukko [54], elective care includes standardized processes with strict pathways, routine processes with predictable outcomes but flexible paths, and non-routine processes where physicians adjust treatment based on individual patient responses. Each subgroup varies in terms of process rigidity and variability.

Mans, Van der Aalst, and Vanwersch [2] noted that, delivering complex care requires coordinated efforts to implement an individualized patient care plan. Non-elective care can be divided into emergency and urgent categories. Emergency care needs immediate action, while urgent care can be temporarily deferred for (say, several days).

2.1.6. Challenge in Healthcare

Healthcare institutions are complex due to the involvement of various stakeholders, such as administrators, physicians, patients, nurses, and different departments [55]. Challenges arise from organizational model failures, discrepancies between business rules and actual processes, and limited interoperability among entities and staff within the organization. Enhanced training and expertise are needed among process specialists in healthcare to promote interdisciplinary collaboration among professionals and prevent varied testing or treatments for similar patient groups with a common diagnosis. Consequently, continuous monitoring is essential to minimize the risks associated with non-compliance violations, medical errors, or harm to patients [56].

According to Yansen Mandacan [47], collaboration in patient diagnosis and treatment presents challenges such as differing routines, knowledge between disciplines, and professional hierarchies. Healthcare professionals need to be on the same page to avoid conflicting advice that could harm a patient's trust. Competition among professionals from different backgrounds may impede goal alignment.

The importance of professional hierarchies and goal alignment in interprofessional collaboration is stated in [47], alongside factors such as structural, psychological, and educational aspects that can facilitate or hinder effective collaboration. Structural factors include physical and organizational settings for collaboration, while psychological factors center on team members' readiness to adjust their work approach. Educational factors pertain to the capacity of team members to engage in collaborative efforts effectively.

Managing clinical pathways is challenging due to the prolonged and uncertain nature of activities, along with the participation of numerous individuals with their own preferences and resources. These factors contribute to inefficiencies such as long wait times, increased costs, deviations from standard procedures, bottlenecks, and frequent interruptions in healthcare operations [41].

2.2. Process Mining

The concept of process mining was initiated from the field of business process management, encompassing individuals, workflows, and technology. It offers a comprehensive understanding of the business process, leading to time savings and enhanced flexibility [57].

Process mining is a method intended to derive process models from event data to make meaningful use of event data. Yansen Mandacan [47], citing Van der Aalst (2016), highlights several advantages of extracting process models from data. Using data for process modeling rather than relying on expert interviews necessitates less input from domain experts, which can be expensive. Moreover, process mining connects process monitoring with diagnosis for redesigning processes, and facilitating Business Process Management (BPM). Process mining is a modern area of study that encompasses various algorithms for recognizing process models from data in the field of information technology [58].

Process models are described as visual representations that depict sequential steps over time using directed graphs. These graphs are useful for evaluating process performance and identifying bottlenecks and deviations by comparing them to an event log [59]. However, gathering process information through interviews and manually creating process models demands significant labor. Fortunately, process mining approaches have been devised to simplify the creation of process models [60].

Van Der Aalst et al. [61] define process mining as tools and methods used to discover, monitor and improve business processes by extracting hidden knowledge from event logs captured and stored in information systems. Event logs comprise sets of traces that represent a case within a process. Each trace is described as a sequence of events associated with the execution of process activities ordered by timestamp and involving various resources, such as humans and machines [41]. In a hospital, information regarding a patient's treatment procedure is recorded in event logs. These logs consist of sequences of activities organized by the patient, including information about resources and the timing of each activity [62]. Event logs for analysis are typically generated within process awareness information systems such as Workflow Management System (WMS) and Enterprise Resource Planning (ERP) [32].

As pointed out in [64], process mining has two key benefits. Firstly, it enables the discovery of the actual operations of people and/or procedures. Secondly, process mining can be used for delta analysis to compare the actual process with an established one, helps identify disparities that can improve the process. Moreover, process mining offers valuable insights into organizational functioning and discrepancies between established processes and can initiate business process reengineering initiatives or customize process-aware information systems.

2.2.1. Types of Process Mining

Process mining consists of three main categories [61], namely discovery, conformance, and enhancement (refer to Figure 2.2).

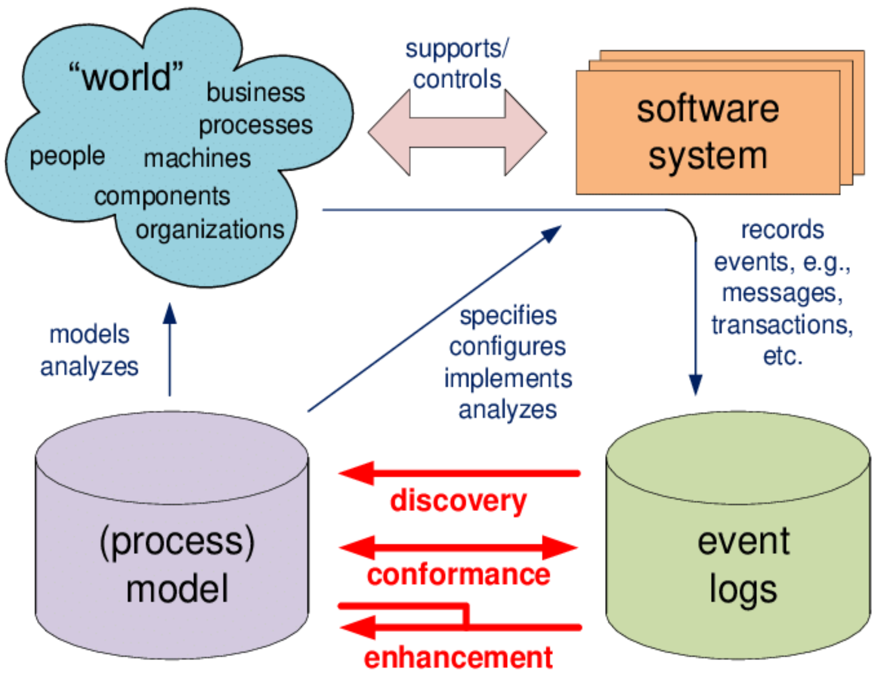


Figure 2.2: The three basic types of process mining [61].

Process Discovery: It refers to a method that involves taking a dataset and generating a process model. Additionally, this approach includes the play-out method, which reverses the process model into a data set and then executes an event log on the process model to display execution times, and frequency of data occurrences, diagnose processes, and identify expectations and recommendations for the process [65]. Discovery involves analyzing the real event log, including case ID, activity name, timestamp, and resource. It allows for generating a new process model using algorithms such as the alpha algorithm, the heuristic algorithm, or the fuzzy miner algorithm. The quality of a good process model is determined by criteria such as fitness, generalization, precision, and simplicity [60].

According to Rashed et al. [41], in common, there is an experiment between the following four value forces dimensions of process model discovery (see Figure 2.3):

- **Fitness**: the discovered model must permit on behalf of the activities noticed in the data set.
- **Precision:** the discovered model must not permit the use of activities, which was entirely dissimilar in the data set.
- **Generalization:** the discovered model must simplify the sample case activities noticed in the data set.
- **Simplicity:** the discovered model should be promising with the plain model.

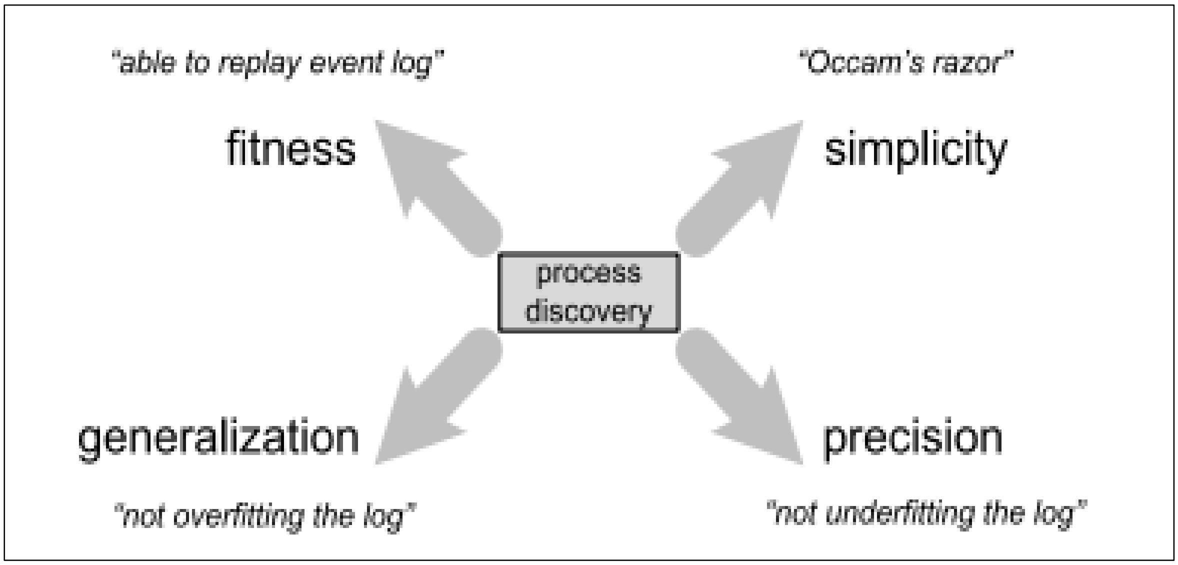


Figure 2.3: Different quality dimensions for process model discovery [60]

Conformance Checking: It involves utilizing a model and an event log as inputs. The model could have been created manually or discovered through process discovery. In conformance checking, the behavior modeled and the behavior observed (i.e., event log) are compared [8]. The model ensures accurate data representation, despite real-time scenarios. Deviations are detected during conformance checks, and replay fitness calculates ideal alignment. Variable expenses for moving activities are considered. Compliance verification confirms dataset adherence to the process model [65]. Process conformance involves replaying log events on a model to identify inconsistencies, providing valuable insights for auditing and compliance purposes [13]. In this stage, it is possible to evaluate the real implementation of the pre-defined process flow, as well as identify bottlenecks, but also find alignments [66]. Assessing conformance involves examining deviations from two perspectives. The first scenario addresses instances where the model is incorrect, while the second pertains to situations where cases deviate from the model, necessitating corrective measures [60]. According to Van Der Aalst [8], conformance checking serves various purposes, such as:

- Assessing the accuracy of documented processes is needed to ensure they reflect reality correctly.
- Identifying commonalities among deviating cases.
- Pinpointing process fragments are where deviations are most frequent.
- Serving auditing purposes.
- Evaluating the quality of a newly discovered process model.
- Directing evolutionary process discovery algorithms.
- Serving as a basis for enhancing models.

Enhancement: Use the data acquired from the real process to expand a process model. The resulting model, designed to answer particular issues, can be a new model entirely or an enhanced version of the original. Enhancement techniques include dotted chart analysis for time analysis and decision mining for extracting branching circumstances [65]. To effectively address problems like inconsistent personnel management or bottlenecks in the organization's processes, improving a process model from various perspectives such as the social network perspective, the time perspective, organization perspective, is necessary to enhance its performance [37].

2.2.2. Process Model and Reality

Making a solid connection between a process model and the real world, as represented by an event log, is a crucial component of process mining [60]. Figure 2.4 provides an illustration of this relationship using the terms play-in, play-out, and replay.

- **Play Out:** Play-out denotes the classical application of process models, applicable for both analyzing and implementing business processes [60]. Here, Kukreja and Batra [57] initiate from a model and produce behavior. The traces could have been acquired by iteratively engaging in the "token game" utilizing the Petri net.
- **Play In:** It indicates constructing models based on data traces [57]. Behavior serves as the input, aiming to develop a model. Play-in, also known as inference, is frequently associated

with this process. Examples of play-in techniques include the α-algorithm and various other process discovery approaches [60].

- **Replay:** Replay employs both an event log and a process model as inputs. The event log is essentially "replayed" on the process model [60]. This facilitates the analysis of the performance and conformance of traces against the model. Additionally, it serves for bottleneck analysis and pinpointing areas for system enhancement [57]. Through the replaying of event logs, predictive models can be constructed [60].

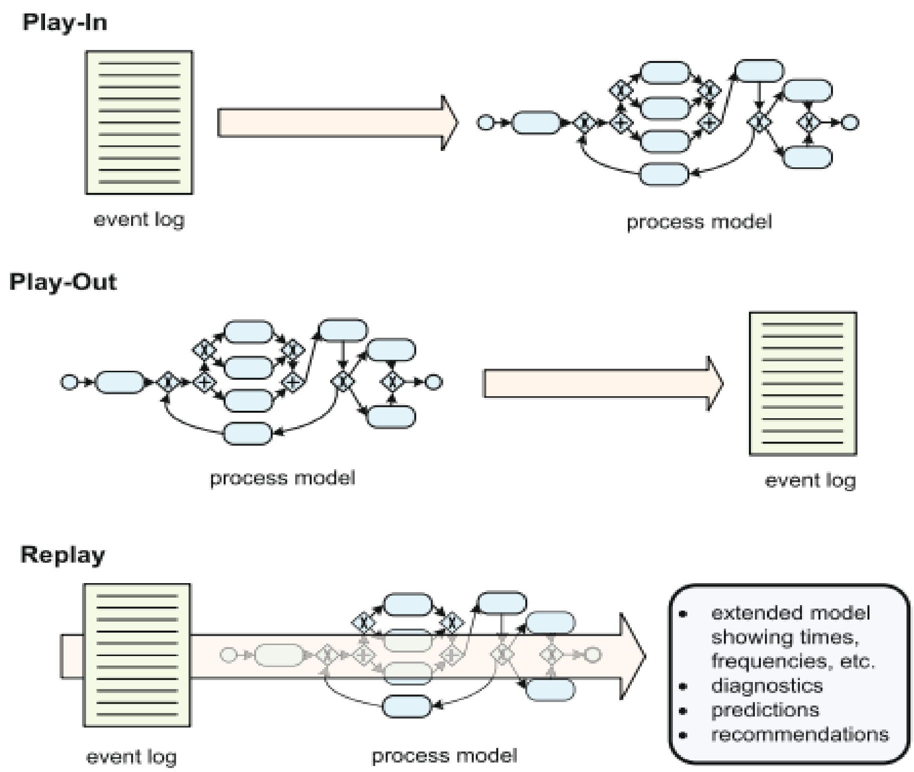


Figure 2.4: Three ways of relating event logs and process models [60]

2.2.3. Process Mining Techniques

Various techniques have been developed for process mining, including the alpha miner, heuristic miner, inductive miner, fuzzy miner, and SNA. According to Van der Aalst [60], the diversity among process mining techniques based on their process discovery methods and handling of representational bias and noise/incompleteness. Representational bias pertains to a modeling language's capability to represent different process structures. Several modeling languages, such as Petri nets, BPMN, causal nets, and process trees, can express and model processes [60]. The second-factor concerns noise and assumptions about incompleteness. Process discovery methods are expected to address noise and incompleteness in event logs. Here, noise refers to rare process instances in event logs that significantly differ from the main process, while incompleteness relates to the ability of a method to discover a generalized model that encompasses all presented instances in an event log, along with other potentially similar processes that may be absent from the log [55].

2.2.3.1. Process Modeling Using Heuristics Miner

Due to the diverse nature of personal activity data, creating a single comprehensive process model to represent all potential individual paths is challenging. To address this, a heuristic-based algorithm can be applied to focus on the most frequent paths taken by each individual. This heuristic mining technique aims to handle the variability and noise inherent in self-tracking logs [23]. By employing heuristic mining, infrequent activities can be filtered out from the dataset, as activities occurring only a few times are unlikely to be relevant to healthcare processes. Additionally, the dataset may contain various attributes, some of which may not be relevant to the process mining task. It is essential to select only the relevant attributes for analysis [47].

The heuristic miner was developed to address noise and incompleteness issues in event logs, with a focus on extracting event relations, such as identifying dependencies between two events. As outlined in [38], constructing a process model using the heuristic miner involves three key steps.

- The initial step involves extracting the following
  - Trace with frequency: it is the number of times of a specific trace occurs.
  - Count order relation: it is the number of times b directly follows a.
  - Dependency measure: it is calculated by the following formula [38].

$$|a=>b| = \frac{(|a>b|-|b>a|)}{(|a>b|+|b>a|+1)} \tag{1}$$

  The sign > implies directly follows, i.e. a>b means a is directly followed by b

- The second phase involves constructing a graph using the extracted dependency matrix. The Equation (1) computes all the cells within the dependency matrix, resulting in values ranging from -1 to +1. A value of 0 indicates no observed relationship between the activities. If a direction is observed, a positive number is assigned, whereas a negative number signifies a counter-directional relationship [41].
- The final step entails designing a process model based on the insights gained from the second step.

The dependency graph reveals the “backbone” of the process model. This backbone is used to discover the detailed split and join behavior of nodes. If an activity has multiple input arcs, then the heuristic miner analyzes the log to see whether the join is an AND-join, an XOR-join or an OR-join. In case of an OR-join, the detailed synchronization behavior is learned. If an activity has multiple output arcs, then the “split behavior” is learned in a similar fashion [8].

2.2.3.2. Process Modeling Using Inductive Miner

The inductive miner algorithm ensures the creation of sound and fitting process models that can reproduce all observed behavior within a finite timeframe. It uses a block-structured framework that guarantees soundness and fitness, supporting the evaluation of various quality criteria and the integration of new blocks or operators without modifying the framework. However, the inductive miner faces challenges with prolonged event dependencies and may produce inaccurate cut-off points in large event logs, potentially resulting in a "flower model" [70].

The algorithm breaks down the task of finding a process model for a log L into subprocesses by splitting LLL using a divide-and-conquer strategy. The method relies on constructing a directly followed graph based on the event log and utilizes this graph to identify different process relationships. This technique involves detecting various cuts within the directly followed graph generated from the event log [70]. These cuts include **Exclusive OR cut (**separates activities into different groups that have no relations between them), **Sequence cut (**shows a "directly follows" relation from the previous group to the next group, but not vice versa), **Parallel cut (**comprises a start activity and an end activity, with activities between the groups having "directly follows" relations), **Redo loop cut (**divides cuts into two parts: a "do" part and a "redo" part. Only actions from the "do" section can start and finish an event log from a redo loop cut). Inductive miner identifies key splits, including sequential, parallel, concurrent, and loop structures, as illustrated in Figure 2.5.

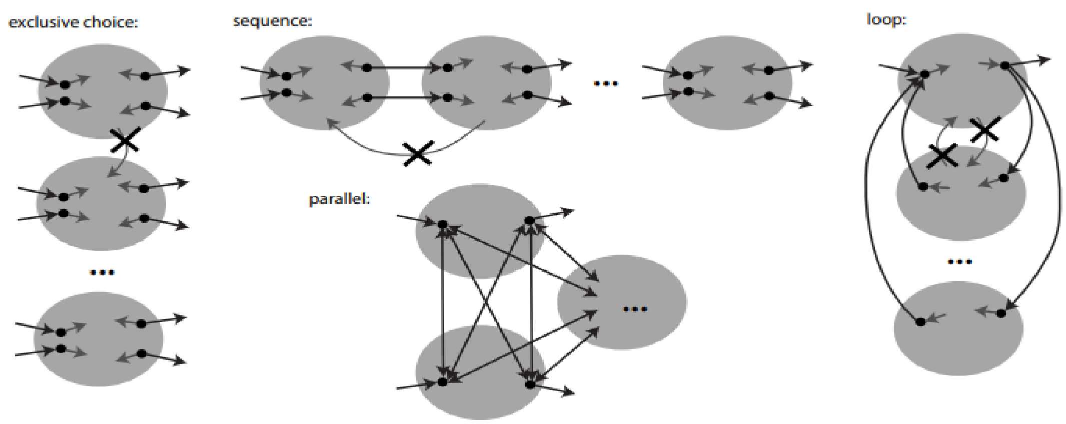


Figure 2.5: Cuts of operators for directly following graph [70]

2.2.3.3. Social Networking Among Activity Performers

Event logs, which include information on individuals carrying out activities, can be used to deduce causal relationships and social interactions by analyzing the sequence of events. This data allows the construction of a sociogram, a graph or matrix representing relationships among activity performers. Sociograms can then be analyzed using Social Network Analysis (SNA) tools to identify cliques (groups of related individuals), measure network density, and evaluate an individual's role, such as centrality or power within the network [73].

In a mathematical sense, a sociogram is a graph (P, R) where P is the set of individuals (in the context of process mining referred to as performers) and R ⊆ P X P. In the event that the graph is weighted, an extra function W is present, which values each element of R. When looking at the graph as a whole there are notions like ***density***: the number of elements in R divided by the maximal number of elements. Another is the ***geodesic distance*** of two nodes: the distance of the shortest path in the graph based on R and W. When looking at one specific individual, many notions can be defined, if all other individuals are in short distance to a given node and all geodesic paths visit this node, clearly the node is very ***central.*** As noted in [63], there are different metrics for this intuitive notion of centrality. The Bavelas–Leavitt index of centrality (i.e., Equation (2)) is a well-known example that is based on the geodesic paths in the graph. Let i be an individual (i.e., $i \in P$) and $D_{j,k}$ the geodesic distance from an individual *j* to an individual *k*.

$$BL(i) = \frac{\left(\sum_{j,k} D_{j,k}\right)}{\left(\sum_{j,k} D_{j,k} + D_{i,k}\right)} \tag{2}$$

Other similar metrics are ***betweenness***, which is a ratio based on the number of geodesic paths that visit a specific node, and ***closeness***, which is 1 divided by the total of all geodesic distances to a certain resource.

Other notions include the following

- the ***emission*** of a resource (i.e., $\sum_j W_{i,j}$)
- the ***reception*** of a resource (i.e., $\sum_j W_{j,i}$)
- the ***determination*** degree (i.e., $\sum_j W_{j,i} - W_{i,j}$)
- ***sociometric status*** which is determined by the sum of input and output relations, (i.e., $\sum_j D_{j,i} + D_{i,j}$)

These metrics allow for the analysis of many facets of an organization's social structure when combined with a visual depiction of the network. To derive meaningful sociograms from event logs, Van Der Aalst, Reijers, and Song [63] identified different metrics. Each metric assigns a weight $W_{i,j}$ to the relationship between individuals i and j. If $W_{i,j}$ is above a certain threshold τ, it will be included in R(i.e., $(i,j) \in R$ if and only if $W_{i,j} > \tau$ for any $i,j \epsilon P$). This way we get a weighted graph (P,R,W) that can be used by process mining tools.

Van Der Aalst, Reijers, and Song [63] focuses on four types of metrics that can be derived from event logs. The first one is metrics based on (possible) causality. It helps to monitor for individual cases how work moves among performers. When there are two following activities, the first of which is accomplished by i and the second by j, there occurs a handover of work from i to individual j within case i.e., process instance). The second metric is based on joint cases. This approach ignores causal dependencies but simply counts how frequently two individuals are performing activities for the same case. The third metric focuses on individual tasks, assuming stronger connections between people doing similar activities rather than collaborative work on shared cases. The last metric is based on special event types. It considers the type of event. Events can include the execution of activities as well as reassigning an activity from one individual to another. From an SNA point of view, these observations are particularly interesting as they represent explicit power relations.

2.2.3.4. Model Quality Evaluation Methods

In this study four quality metrics, such as fitness, precision, simplicity, and generalization are used to evaluate the performance of the discovered process models.

- **Fitness metri**c: This study used the ProM plugin named "Replay a log on Petri net for conformance analysis" which was derived from the alignment-based conformance checking in [67]. This method aligns traces from an event log with a Petri net to map activities to transactions, which is key for measuring conformance. It highlights deviations between the trace and the net, using a severity cost function to quantify the fitness of the alignment based on the total severity cost of deviations. The absolute fitness value proposed in [67] can be higher than 1. Thus, a relative fitness metric is proposed, that always provides values between 0 and 1. In practice, it is often desirable to have a quantification of fitness in the range of 0 (very poor fitness) to 1 (perfectly fitting). The relative fitness between event log and Net using oracle function orc^{sc} and severity function sc is (i.e., Equation (3)):

$$rfit(L, N, orc^{sc}, sc) = 1 - (\frac{1}{|L|} . \sum_{\sigma \varepsilon L} \sum_{\gamma \epsilon \Gamma \sigma, N} (orc^{sc}\ (\sigma)(\gamma). \frac{\sum_{(x,y)\epsilon\gamma} sc\big((x,y)\big)}{lim(\sigma, N, sc)})) \tag{3}$$

  Where A $\subseteq A$ be a set of activities. Let L$\epsilon B(A^*)$ be an event log over A; N = (P, T, F, α, m_i , m_f) be a Petri net over A; $sc: (A^{\gg} x T^{\gg}) \rightarrow IR$ be a severity cost function; $orc: \{\sigma \in L\} \rightarrow ((A^{\gg} x T^{\gg})^* \rightarrow IP$ be an optimal oracle function of L and N.

- **The precision and generalization metrics** are assessed using the "Measure the Precision/Generalization" plugin, which requires three inputs: The Petri net, the event log, and the fitness result. This plugin evaluates precision based on the Equation (4) introduced in [68]. It measures precision by aligning the event logs to the process model. If all activities predicted by the model are actually observed in the logs, the precision score is '1'.

$$Precision(Log, Model) = \frac{1}{|\varepsilon|} + \sum_{e \epsilon \varepsilon} \frac{|en_L(e)|}{|en_M(e)|} \tag{4}$$

  Where each event $\boldsymbol{e \epsilon \varepsilon}, \boldsymbol{\epsilon}$ is the collection of unique events, $en_M(e)$ the number of enabled activities in the model, and $en_L(e)$ is the number of observed activities actually executed. The precision value ranges between '0' and '1'. If the value is close to '0', the model is under fitting, while a value near '1' indicates a more precise model. Similarly, the "Measure the

Precision/Generalization" plugin evaluates generalization using an approach similar to that used for measuring precision. The approach based on the Equation (5) to quantify the generalization:

$$Generalization(Log, Model) = 1 - \frac{1}{|\varepsilon|} + \sum_{e \epsilon \varepsilon} pnew(|diff(e)|, |sim(e)|) \tag{5}$$

Where each event $e\epsilon\varepsilon, \varepsilon$ represents a set of distinct events,, $pnew(w; n)$ is the estimated probability that the next visit to the state $stateM(e)$ will reveal a new path not seen before, $w = |diff(e)|$ denotes the count of unique activities observed leaving the state s and n = $|sim(e)|$ is indicates how many times s was visited by the event log.

- **The simplicity metric** evaluates how simple the process model is to understand for a human. As a result, this dimension can take the process model into consideration independently of the observed behavior. Since there are different ways to describe the same behavior using different process models, choosing the simplest one is best. Several measures exist to measure how simple a process model is, therefore, this study adopts the approach proposed in [41], comprising two steps. Firstly, validating the soundness of each model generated by the mining algorithms. The second step, assessing control flow complexity metrics, density, cyclomatic complexity, and structuredness.

Check the soundness: A process model is sound when it meets three essential criteria. The first is the "option to finish," meaning that the process can proceeds from any state to the process model final state. The second criterion is "proper completion," which occurs when the process reaches its final state with no token is left behind. The third criterion is the absence of " no dead transition", ensuring every transition in the model is enabled. The study employed the "Analyze with Woflan" ProM plugin to verify the soundness of every model generated by the discovery miner algorithm.

Measuring the metrics of density, Cardoso, cyclomatic, and structuredness: the density is the relation between the number of arcs and the maximum number of arcs between all nodes, Cardoso is the sum of all process options based on the number of splits of each type and its number of outgoing arcs. Cyclomatic is the number of nodes within cycles with regard to the all number of nodes, it is calculated by the rule: $CM = |E||V| + P$; where E is the number of nodes, V is the number of edges, P is several connected components. Structuredness in a process model is the percentage of well-structured components compared to unstructured ones. It is measured by the ratio of nodes in the reduced process graph to those in the original process graph. Cardoso, Cyclomatic, and Structuredness metrics were obtained using the ProM plugin "Show Petri net metrics" in the study.

2.2.4. Process Mining Perspectives

Process mining analyzes event logs to uncover inefficiencies, deviations, and collaboration patterns in processes, offering insights for optimizing performance and enhancing efficiency [31]. Process mining techniques enable analysis from four different perspectives, such as control flow, organizational, performance, and case [13]. These perspectives provide distinct views of process data, such as activity sequences, resource allocation, individual cases, and timing. The three process mining approach categories are orthogonal to these perspectives, meaning they focus on objectives like process model discovery, compliance verification, and process improvement, each through a different analytical lens.

- The **control flow perspective** focuses on the ordering of activities. It helps identify patterns, such as loops, branches, and sequences of activities and can be used to detect potential issues, such as missing activities or unnecessary steps [13]. Given an event log containing a set of traces, a suitable process model describing the behavior seen in the log will be automatically constructed. In particular, the process model describes the causal dependencies between activities, focusing on their orderings [60]. The control flow perspective helps organizations identify deviations from the intended process, improving compliance and reducing errors. It also reveals potential optimizations by analyzing activity sequences and execution times, enabling process redesigns that enhance efficiency, lower costs, and boost customer satisfaction [60].
- **Performance Perspective** quantitatively evaluates a business process effectiveness, efficiency, and key performance indicators, helping organizations benchmark against industry standards and identify areas for improvement through event log and process model analysis [60]. Key time measures in process analysis include average activity duration, longest completion time at specific points, and total end-to-end duration of each process instance. The control flow and performance perspectives are closely linked, often starting with control flow analysis, followed by performance analysis, with results displayed on the control flow graph.

This approach visualizes event timelines, enabling comparisons across different process instances [74]. According to Van der Aalst [60], the time dimension of performance includes four key indicators. The first one is a lead time (total duration of a case), The second is a service time (active working time on a case). The other is waiting time (idle time while waiting for resources). Lastly, synchronization time (inactive time awaiting external triggers or parallel branch synchronization).

- **Organizational Perspective** analyzes resource data in event logs to understand the roles and functions of both human and non-human elements in process execution [13]. It involves two models. The one is organizational model, showing the organization's structure, and the other is the social network model, illustrating communication dynamics [75]. This analysis reveals details about who performed specific events, offering insights into roles, organizational divisions, and relationships, helping to organize the framework or visualize social network dynamics [82]. Resource information can enhance process analysis, resource allocation, and organizational pattern analysis by revealing workflow patterns and relationships between resources and activities [35]. Social Network Analysis (SNA) provides insights into cross-departmental collaboration by examining interactions, focusing on coordination, communication, and shared information [35]. While traditionally based on interviews and questionnaires, modern SNA now leverages extensive electronic data for analysis [60].
- **Case Perspective** emphasizes individual case characteristics, including attributes, event features, and performance measures, to identify patterns influencing decisions through methods like decision tree learning. This approach uses predictor variables to interpret outcomes and, in cases with loops or insufficient data, applies probabilistic methods to decision points. The results enhance the process model by incorporating case-specific insights [60].

2.3. Process Mining Methodologies

Several data mining process models have been developed to outline the life cycle of a standard data mining project and research [77]. According to Van Der Aalst et al. [61], existing data mining techniques offer limited support and are not well suited for process mining projects. As process mining gains research focus, new methodologies have been developed to provide more accurate analysis and decision-making tailored to specific industry settings. The following section outlines some of the most widely used process models in process mining.

2.3.1. L∗ Life-Cycle Model

As shown in Figure 2.6, the five-stage L∗ life-cycle model outlines the phases of a process mining project, based on practical experience from over 100 organizations, designed to enhance structured processes like lasagna preparation [77].

The L∗ life-cycle model for process mining encompasses five distinct stages. Stage 0, Initial Planning and Justification, which involves meticulous project planning, including scheduling, resource management, milestone definition, and progress monitoring. In Stage 1, Data Extraction, focusing on understanding and acquiring relevant data, identifying key questions, and gathering historical data and models. Stage 2, Control Flow Model Development and Event Log Integration, where techniques such as α-algorithm, heuristic mining, fuzzy mining, and genetic mining are used to uncover control flow patterns. Stage 3, Integrated Process Model Development, which enriches the control flow model by adding perspectives like organizational, case, and time to address specific inquiries and guide actions. Finally, Stage 4, operation support, leverages current data for detection, prediction, and providing actionable recommendations to users.

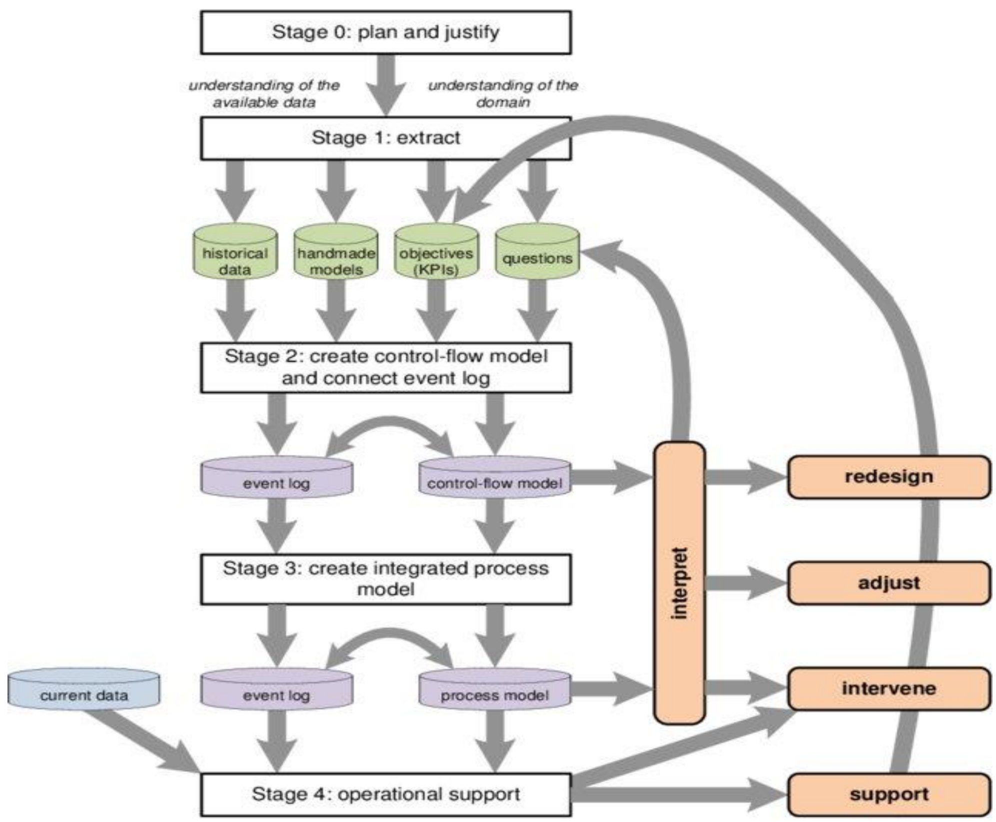


Figure 2.6: The L∗ life-cycle model [77].

2.3.2. PM2 Methodology

The Process Mining Project Methodology (PM2) is tailored to support projects seeking to enhance process performance or compliance. It encompasses diverse process mining and analysis techniques suitable for structured and unstructured processes alike. The PM2 facilitates swift analysis iterations and evolving insights, taking into account established best practices. The PM2 methodology comprises six stages (see Figure 2.7), each involving different input and output objects [78].

Planning initiates the project by identifying research questions, selecting business processes, and forming the project team. This groundwork leads to the Extraction stage, where the project scope is determined, event data is extracted, process knowledge is transferred, and information systems are utilized. Following this, Data Processing retrieves event data and develops process models, incorporating insights from the research questions and business process systems. The project then moves to Mining and Analysis, where process mining techniques analyze event logs to evaluate performance and compliance, using existing models for conformance checking and generating findings. These findings are then addressed in the Evaluation stage, which links them to project goals, producing ideas for improvement and new research questions. Finally, the Process Improvement and Support stage applies evaluation insights to implement modifications and support ongoing operations, ensuring continuous enhancement and optimization.

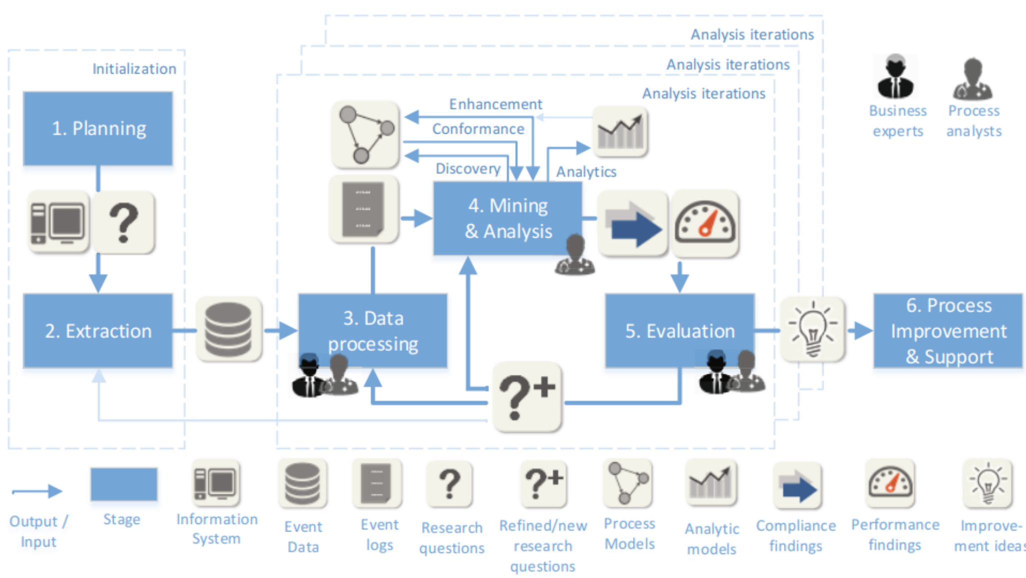


Figure 2.7: An overview of the PM² methodology [78].

2.3.3. Process Diagnostics Method

The methodology, which was presented by [31], attempts to provide a concise, comprehensive overview of the information systems operations. There are five phases to it. The methodology is depicted as a process in Figure 2.8. The first two phases serve as input for control flow analysis as well as role analysis; in the meantime, control flow analysis is also a source of information for performance analysis.

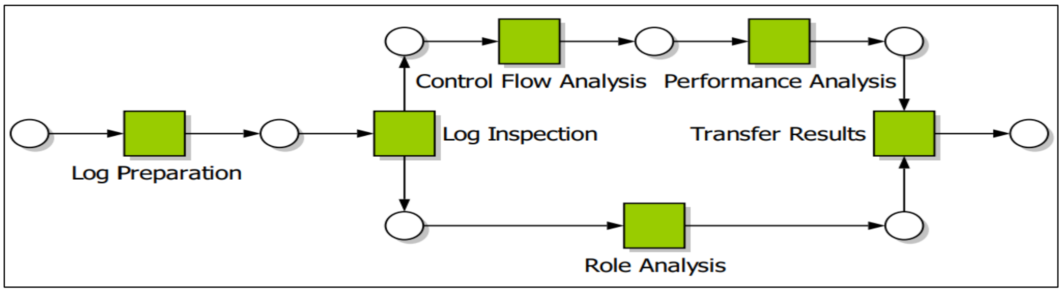


Figure 2.8: Phases of the methodology [31].

The first stage in process mining analysis is Log Preparation, which involves preprocessing the event log by identifying events, adjusting the log format, and clarifying timestamps. Log Inspection then assesses log statistics such as case numbers, event categories, and occurrences. Control Flow Analysis examines the actual process flow, performing conformance checks or using algorithms to identify control flow patterns. Performance Analysis identifies bottlenecks and computes throughput times using log replay and dot chart analysis. Role Analysis addresses questions about who performs which activities and their collaborations, using tools like the role-activity matrix to understand task distribution. Finally, Transferring the Results involves discussing findings with the client to ensure understanding and providing guidance for optimizing the information system.

2.3.4. BPA Methodology

The methodology of business process analysis (BPA) in the healthcare industry builds upon the work of [31], primarily emphasizing methods that are capable of handling the characteristics of healthcare processes to investigate rare behaviors and variations in processes. To do this, this methodology adds a new step following log inspection, as illustrated in Figure 2.9, to the work of [31]. This additional stage includes pre-process analysis and log clustering algorithms. Running a sequence clustering algorithm, creating a cluster analysis diagram, comprehending regular process behavior, spotting process variants and irregular behavior, carrying out hierarchical sequence clustering if required, and choosing the most interesting clusters for additional investigation are all included in this process [3].

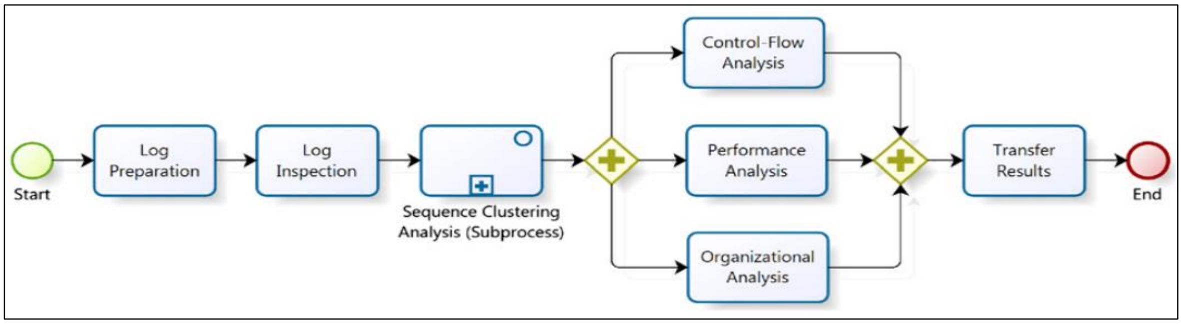


Figure 2.9: Methodology for BPA in healthcare [3].

2.4. Process Mining in Healthcare

Healthcare processes involve a series of activities to diagnose, treat, and prevent diseases to improve patient health. These activities are carried out by various types of medical experts (such as physicians, nurses, technical specialists, dentists, and clerks) and may differ between organizations [2]. The healthcare domain undergoes a lot of dynamic, inaccurate, and complex data that needs to be processed to provide welfare to the hospital team [57]. The healthcare sector is also confronted with new patients at various condition stages and new cases daily, representing the dynamic nature and complexity of healthcare processes performed [1]. The increase in demand for medical services due to population aging and improved standards of living has led to high-quality consultations from a medical progress perspective and an optimal clinical process from a business process perspective being provided to patients [32].

Improving healthcare processes can significantly enhance patients' quality of life. To achieve these goals, analyzing both clinical and organizational processes is essential [1]. Traditionally, methods like business process redesign, evidence-based medicine, and lean were used, relying on manually recorded data. However, with the adoption of information systems such as Practice Management Systems (PMS), Electronic Medical Records (EMR), Computerized Physician Order Entry (CPOE), and Picture Archiving and Communication Systems (PACS), healthcare institutions now manage and analyze data more efficiently [32]. These systems have paved the way for the emerging field of process mining, which helps improve clinical processes and patient management while reducing costs [6].

Process mining in healthcare can improve service quality, enhance collaboration, reduce waiting times, and offer additional benefits such as understanding resource and patient behavior, suggesting process redesigns, analyzing performance, and identifying bottleneck-causing activities [6]. It allows health experts to understand the actual execution of processes by discovering models, checking conformance with guidelines, and finding improvement opportunities [1].

The analysis of process mining techniques in healthcare reveals five key inquiries: commonly followed paths, variances in clinical pathways, adherence to guidelines, identified bottlenecks, and internal relationships [66]. These insights create a framework for utilizing process mining. Munoz-Gama et al. [11] highlight that process mining can help answer critical questions about patient flow differences, adherence to care pathways, bottlenecks in the healthcare process, and interactions among clinical experts.

2.4.1. Application of Process Mining in Healthcare

Reviews of past research and current trends in the field by De Roock and Martin [79] categorized processes into clinical and organizational types, with clinical processes being the most frequently discussed topic which accounts for 63.9%. The majority of studies focused on discovery (81%), followed by conformance (31.2%) and enhancement (14.8%), often in combination. It examines perspectives like control flow (68.1%), time (27%), case (10.3%), and organizational process (9.1%). The paper also distinguishes between the development of new algorithms and the application of existing ones, with the latter being more common (59.3%). Popular algorithms include the heuristics miner, fuzzy miner, and inductive miner. Additionally, process mining is increasingly being combined with other techniques, such as clustering and machine learning.

The literature review by Batista and Solanas [80], categorized 55 articles across nine dimensions to highlight their contributions from various perspectives. These dimensions include the objectives of the process mining analysis, types of medical facilities, and medical data preprocessing techniques. Regarding process mining analysis objectives, 10% of the papers focus on validation, 27% on exploration, 29% on comparison, 8% on KPI, 4% on visualization, and another 4% on state-of-the-art approaches, with the remaining 14% addressing theoretical aspects. In terms of medical facilities, 56% of the articles use data from hospitals worldwide, 31% from emergency units, and 13% from other sources. For preprocessing techniques, 22% of the studies apply noise filtering, 25% use aggregation and generalization, 28% employ standardization and semantics, and 25% focus on privacy protection.

The review conducted by Rojas et al [1] identified and classified common aspects across reviewed articles. ProM is the most frequently used tool, featured in 42% (31 out of 74) of the case studies, while Disco is used in 10% (8 cases). The study categorizes implementation strategies into three types: direct implementation, which directly uses data from HIS sources to create event logs and is the most prevalent; semi-automated, which involves custom developments for extracting and constructing event logs by linking multiple data sources; and integrated suite, which connects data sources and applies process mining techniques without requiring extensive technical knowledge, used in 9% (7 cases). Geographically, the majority of case studies are in Europe (73%), with fewer studies in Australia, Asia, and North America, and none in Africa or Latin America. The Netherlands has the highest number of case studies, followed by Germany and Belgium.

2.4.2. Challenges in Applying Process Mining in Healthcare

Process mining in healthcare faces several significant challenges, primarily centered around data quality, process complexity, technique diversity, and evolving methodologies.

Data Quality and Completeness: One of the foremost challenges is ensuring the quality of data used for process mining [66]. Healthcare data often suffers from incompleteness, outliers, and varying levels of granularity due to its distributed nature across multiple sources [8]. These issues complicate the creation of comprehensive event logs, making data modeling, extraction, integration, and visualization particularly challenging [81]. Establishing representative benchmarks, such as datasets and quality criteria, is essential to compare and enhance tools and algorithms [8].

Process Complexity and Abstraction: Healthcare processes are inherently complex due to the diverse needs of patients and the intricate connections between numerous events. Simplifying these complex models requires a deep understanding of both medical and non-medical factors. The abstraction stage, which involves domain experts, is resource-intensive and difficult to execute effectively. To address this, machine learning techniques like Hidden Markov Models and Expectation-Maximization have been proposed to aid in abstraction. Additionally, visualizing these abstracted models remains challenging, necessitating new methods for transparent and understandable model visualization [55].

Handling Concept Drift and Process Evolution: Concept drift, where healthcare processes change over time during analysis, presents another challenge. Process mining techniques must be adaptable to these changes, requiring tailored methodologies and frameworks to ensure accurate analysis. This involves exploring novel conformance checking and enhancement techniques specifically designed for healthcare, as well as maintaining a focus on real-life data to ensure that findings are relevant and applicable [8] [11].

Technique Diversity and Model Selection: The diversity of process mining techniques and software poses challenges in terms of evaluating their quality and effectiveness [81]. Traditional metrics like fitness and precision may not be adequate for complex healthcare processes. There is a need for independent evidence or comparisons of these techniques, along with the creation of representative benchmarks to guide practitioners [8]. Selecting the most suitable model for complex healthcare processes is crucial, and criteria based on state importance rather than traditional metrics are suggested to improve model selection [55].

Integration, Usability, and Privacy: Integrating process mining insights with Healthcare Information Systems (HISs) and providing operational assistance in real-time are essential for enhancing the usability and effectiveness of process mining in healthcare. However, ensuring that these techniques are accessible and understandable to non-experts, such as healthcare professionals, remains a priority [8]. Safeguarding patient privacy and data security is crucial in process mining, particularly when handling sensitive healthcare information. It is essential to understand healthcare processes from the patient's perspective and to continuously evolve process mining techniques in line with advancements in the healthcare field to ensure their ongoing relevance and effectiveness [11].

2.5. Related Works

The importance of process mining in various industries, including healthcare, has led researchers to explore its application in the healthcare system. In this context, a review of relevant research studies is presented below, starting with the most recent ones and progressing to older works. The researcher then aimed to identify research gaps (see Table 2.1) that can be addressed via the current research.

Rashed et al. [41] proposed utilizing process mining methods in healthcare to gain insights into care processes, using data from a cardiac surgery unit in an Egyptian hospital. Various popular discovery mining algorithms were applied to discover the process model, which was then analyzed from control flow, performance, and organizational perspectives. The study's approach comprised three main phases: data preprocessing, model discovery, and analysis, all executed within the ProM framework with its extensive plugins. Clustering techniques such as ActiTrac and Markov were employed, resulting in four and two clusters, respectively, though the latter was refined based on input from hospital experts. The refined clusters were based on the main activities designated by cardiac consultants, such as different types of surgeries and medication-related activities, and were converted into petri net models for further analysis. Quality metrics were analyzed to determine the most suitable process model for describing care flows across all patient groups, considering fitness, precision, generalization, and simplicity. Simplicity was assessed by checking model soundness and measuring control flow complexity using Cardoso, cyclomatic, and structuredness metrics. The inductive miner algorithm emerged with top scores for fitness (94.77%) and generalization (99%), closely followed by the ETM miner (87% and 99%, respectively). Conformance checking indicated a 95% compliance between the base model and event logs from the information system, highlighting areas for improvement in certain hospital services according to global statistics. An analysis of conformance checking using the inductive visual miner revealed numerous skipped activities as per the model. Additionally, assessments using dotted charts provided insights into patient cases and throughput times for different clusters. The study recommends focusing on enhancing business process performance by improving prediction models and recommendation systems through a combination of process mining and machine learning techniques.

A case study conducted by Agostinelli et al [82] in a hospital in Rome, Italy, utilizing process mining techniques to analyze patient care flows from event logs recorded by hospital information systems. The study employed the PM2 approach and various plugins for data preprocessing and analysis, including filters like Simple LOG and Time-based LOG, along with tools like explore event log, log visualizer, inductive visual miner, transition system miner, social network miner, and dotted chart. three key questions were identified by hospital management for investigation: the analysis of outpatient clinic services provided by the radiology department, the temporal distribution of patient abandonments from the emergency room, and the performance comparison between newly introduced and pre-existing operative blocks. The findings revealed instances of repeated services for single patients within the radiology sub-department, many-to-many relationships between resources, and consistent effort levels for healthcare services provided without reservations. Regarding emergency room activities, the inductive visual miner plugin highlighted cases where patients left before or during medical examinations, totaling 857 patients with varying waiting times before abandonment. In hospitalization activities, it was observed that four rooms in the new block became fully operational in August 2017 while others remained available for exceptional needs as confirmed through SNA due to the lack of interaction between rooms in each block. The hospital management expressed interest in further economic analysis, planning to integrate new data with clinical datasets containing economic values associated with each surgical intervention. The aim is to define revenue concepts associated with each patient classified by Diagnosis Related Group Code (DRG), representing a classification instrument for the hospital's overall work for remuneration purposes. This approach aims to provide insights into the economic aspects of hospital operations, facilitating informed decision-making and resource allocation strategies to optimize revenue generation and operational efficiency.

The study in [43], demonstrated the use of process mining to analyze patient treatment procedures in a cardiology hospital. The evaluation of system performance was based on event logs from daily hospital processes. Various activities such as patient reception, medication booking, ID generation, symptom-based redirection, pharmacy queues, and waiting area guidance were identified. Data collection included interviews with medical professionals and obtaining event logs from the hospital's information system. The paper emphasized the importance of sustainable resource utilization in healthcare for maximizing benefits and reducing costs but also pointed out challenges like limited information systems in healthcare organizations. Overcoming this challenge involves storing comprehensive patient digital footprints and enhancing data science and process mining expertise. Additionally, the authors expressed interest in exploring applications of process mining in agriculture and education to enhance understanding and improve processes.

The application of ProM tools and four key process mining algorithms was highlighted by Tibeme, Shahriar and Zhang [50], which demonstrated how ProM can address questions regarding control flow (how), organizational roles (who), and case details (what). The study centered on four main algorithms, testing them on an artificially created clinical dataset designed for this purpose. This dataset, in Excel format, included attributes like case ID, patient name, event, and completion time, encompassing around 150 individuals. The Excel data was imported into ProM and transformed into the XES file format. Subsequently, process models were derived using the alpha miner, heuristic miner, inductive miner, and fuzzy miner algorithms. the paper discusses the evaluation and detailed results of applying these algorithms. The study found that while the alpha miner effectively represents process flows, it does not identify bottlenecks within the process. The heuristic miner improves upon the alpha miner by incorporating frequency data and illustrating the relationships between activities. The inductive miner further enhances these by managing large and noisy event logs to produce robust process models. The fuzzy miner generates simplified fuzzy models that depict specific views of a process. The paper concludes that process mining techniques are effective for structured processes, and future research will aim to extend these techniques to unstructured processes.

Dogan [84] using a case study analyzed the check-up process through process mining. Data was exported from Microsoft access to Excel and then imported into the process mining tool Disco. The event log inspected in the disco included 2,372 events, 732 cases, and 9 distinct activities. After the 'Patient Record' activity, the process diverges into three paths: in 90 cases, 'MR, BT, USG' was performed; 'Bloodletting' was performed 200 times directly after 'Patient Record', with the process returning to 'Patient Record' 176 times; and 203 cases involved patients exiting the check-up system directly. 'Patient Record' was the most frequently executed activity, occurring 732 times, as it involved both saving patient details and serving as an information desk. The study identifies 'Bloodletting' as the central activity requiring frequent adjustments and re-evaluation due to inefficiency. It is important to analyze the reasons behind these continual adjustments to enhance the process. The authors suggest that patient unawareness of examination procedures could be a factor and recommend updating examination guidelines or providing additional training. To improve the efficiency of bloodletting, addressing these frequent adjustments is crucial. Out of 630 cases, 192 check-up reports were completed, while 203 were halted early. The study reveals that the average check-up duration is 59.5 hours, based on 2,372 events across 630 cases. For future research, the authors recommend using more event log filters to gain detailed insights and specific knowledge about the process.

The case study by Toth et al [59], identified several key challenges in healthcare process mining. One major issue is that healthcare event data is often stored in separate databases, leading to common relational database problems like data redundancy and inconsistency. Another issue is that code systems can be imprecise, with a single code potentially representing a wide range of disease severity or multiple treatments. Additionally, the heterogeneous nature of patient medical care complicates process analysis due to varying conditions. To address these problems, the authors suggest several solutions. For data redundancy and inconsistency, they recommend professional review and manual correction of the data. To tackle the code system issue, they propose reducing the number of codes through aggregation. Finally, they suggest that accurately describing patient cohorts is the most effective way to mitigate heterogeneity in healthcare processes. These measures aim to create more precise process models and effectively manage the challenges in healthcare process mining. The study was conducted to analyze the medical care of colorectal cancer patients using data from the financial database of the Hungarian healthcare sector, covering diagnoses and treatments from January 1, 2009, to December 31, 2014. The data was preprocessed and converted to XES format for use with ProM process mining software. Initially, the generated process models were complex and "Spaghetti-like," so a treatment hierarchy was established to simplify them. The event log included 25,192 cases and 405,862 events. An inductive visual miner was used to discover the process model, which identified five different activities and their frequencies. Diagnostics and chemotherapy were the most frequent activities, while radiotherapy was the least common treatment.

The study presented in [20], developed a methodology for analyzing patient flow complexity within the emergency department (ED) of Acute hospitals in Ireland. Using patient event logs, the paper employs process mining techniques to uncover actual patient pathways, understand the high variability in these pathways among different patient groups, and gain insights into bottlenecks and resource utilization. A real-time patient tracking information system was utilized to monitor patients' journeys within the ED. Due to the unstructured nature of ED processes, the fuzzy miner was used to map out patient flows and assess flow variability. By analyzing event timestamps, the study identified and examined ED performance and bottlenecks. Resource requirements were also analyzed, providing ED managers with insights into actual staff and resource allocation and highlighting gaps between best practices and actual performance. This work aims to reduce decision-making latency in complex healthcare systems. The study continued with the partner hospital to integrate the process mining engine with the Hospital Information System, enabling real-time ED process tracking and further reducing decision-making latency.

The study Yoo et al [34], utilized process mining technology on EHR log data from outpatient care to analyze the impact of environmental changes, such as the construction of a new building, on hospital processes. The study compared data from before and after the changes to measure the effects on consultation wait times, task durations, and outpatient care processes. Key performance indicators, including the total time for outpatient care, consultation wait time, and test wait time, were analyzed to evaluate operational efficiency. Following the new building's establishment, the number of outpatients increased significantly by 154.6% in the cancer center and 67.8% in the clinical neuroscience center. Despite the higher patient volume, consultation wait times decreased by 4.6% in the cancer center and 12.4% in the clinical neuroscience center. Test wait times increased by 62.7% (7.49 minutes) in the cancer center but decreased by 7.8% (0.56 minutes) in the clinical neuroscience center. Additionally, the number of tests per patient rose by 4.4% in the cancer center and decreased by 9.3% in the clinical neuroscience center. The results indicate that the outpatient clinic's operation improved following the environmental changes, with process mining techniques identifying further process improvements. The study confirmed the effectiveness of process mining technology in a hospital setting. Future research will aim to expand the application of process mining technology to a broader range of process change scenarios.

A case study conducted in [83] demonstrated the application of process mining techniques in the urology department of Isala Hospital in the Netherlands to align data with clinical guidelines. The study utilized process mining to analyze deviations, check conformance, and enhance models with conformance-related diagnostics. This approach proved feasible, with tools like ProM and Declare providing valuable insights into process conformance. Moreover, the study applied process mining technology to EHR log data from outpatient care to examine process changes due to environmental factors like the construction of a new hospital building. It assessed the impact of these changes on consultation wait times, task durations, and overall outpatient care processes. The findings indicated that the outpatient clinic's operations improved following environmental changes, with process mining identifying process enhancements. This study validates the effectiveness of process mining technology in a real hospital setting, suggesting its potential for broader application in diverse process change scenarios in future research.

Mans et al [19], conducted a case study to investigate utilization of process mining was demonstrated to analyze the gynecological oncology process at a Dutch hospital, employing event logs extracted from the hospital's information system and processed using the ProM framework. The study investigated the healthcare process for 627 gynecological oncology patients treated between 2005 and 2006, emphasizing care paths and departmental interactions while excluding low-level lab activities. The authors presented findings from a comprehensive analysis of the hospital's event log, focusing on control flow, organizational structure, and performance aspects. Control flow analysis utilized the heuristic mining algorithm to highlight main process flows rather than minutiae. Clustering techniques segmented the process log into nine clusters, with one containing 352 cases sharing similar properties. Social network mining revealed collaborations between hospital departments, particularly with the general clinical chemical lab. Performance insights, presented via dotted charts, indicated more diagnostics and treatment events early in the process, with prolonged instances mainly consisting of regular consultation activities. Future research aims to develop novel mining techniques and creatively utilize existing ones to derive concise, high-level information instead of intricate models showcasing every detail.

Here under summary of related works that applied process mining in healthcare sector are presented in Table 2.1.

Table 2.1: Summary of related works

Author	Problem	Approach	Results	Gap
Rashed et al. [41]	• Analysis of the patients' care flows using process mining	• Case study • ActiTrac and Markov clustering • heuristic, inductive, ILR, ETM, dotted chart	• The inductive miner algorithm achieves the top scores for fitness 94.77% • Compliance between the base model and the event logs is 95%	• Based on the finding, the solution to improve healthcare service not provided
Agostinelli et al. [82]	• Supporting Governance in Healthcare Through Process Mining	• Case study • Simple LOG filter, Inductive Mine, SNA, Dotted Chart	• Repeated services for single patients • Many-to-many relationships between resources, with consistent effort levels despite these complexities • Potential inefficiencies and underutilization in room allocation	• datasets happened in the span of two months
Toth et al. [59]	• Applicability of Process Mining in the exploration of Healthcare Sequences	• case study • An inductive visual mining	• Difficulties of healthcare process mining identified: data redundancy and inconsistency, inexactness, heterogeneous in medical care • The model shows five different activities (pathology, diagnostics, surgery, chemotherapy, and radiotherapy.)	• Evaluation of the result is not done • not analyzed from different perspective

Abo-Hamad [20]	• Patient Pathways Analysis Using Process Mining	• Case study • Fuzzy miner is used to discover the ED patient flow	• Average length of stay is 9.1 hour, 3 h above national target (both admitted and discharge) • Triage activity should take place in the triage room by a registered nurse (RGN). However, the analysis reveals that 68% this activity takes place in the triage room	• Only one model discovery technique is applied
Hakim [43]	• Improving Healthcare Systems through Process Mining	• Case study • DISCO tool • Interviewing doctors and other staff	• sustainable resource utilization maximizing benefits and reducing costs • pointed out challenges like limited information systems in healthcare organizations	• Not analyzed from different perspective such as organizational and performance perspective
Dogan [84]	• Process mining for check-up process analysis	• Case study • Disco tool	• Examined 9 different activities. Bloodletting is the mostly occurred activity • The average duration of a check-up is 59.5 hours	• Discovered model was not analyzed from organizational perspective
Tibeme et al., [50]	• Process mining algorithms for clinical workflow analysis	• Case study • alpha, heuristic, Inductive, and Fuzzy Miners	• Inductive miner is an improvement on alpha and heuristic miner to handle large and noisy event logs, and produce sound process model • Fuzzy miner generates fuzzy models that represent a specific view of a process (for simplification)	• The dataset was only structured data
Yoo et al. [34]	• Assessment of hospital processes using a process mining technique	• Case study • analyze process changes based on changes in the hospital environment	• By construction of a new building, total time spent in outpatient care did not increase significantly, considering that the number of outpatients increased by 154.6% in the cancer center and 67.8% in the clinical neuroscience center, and the consultation wait time decreased by 4.6% in the cancer center and by 12.4% in the clinical neuroscience center	• The treatment time could not be analyzed since the system records only the start time of the treatment
Rovani et al. [83]	• To analyze process changes based on changes in the hospital environment	• Case study • Declare Replayer and the Declare Diagnoser •	• The operation of the outpatient clinic was effective after changes were implemented in the hospital environment • They further identified improvements in processes using the process mining technique	• Only control flow perspective is analyzed • the methodology relies on event logs that record case executions
Mans et al. [19]	• To obtain meaningful knowledge about typical	• Case study • heuristic, clustering, social	• Nine clusters are obtained from the log, the paper showed only the result for the biggest cluster containing 352 cases	• Applied only on structured data • The result of the techniques is

	paths followed by particular groups of patients	network, dotted chart	• The mining result shows that the general clinical chemical lab is highly involved in the process and interacts with many departments	"spaghetti-like" models showing all the details.

From Table 2.1, it can be observed that a significant amount of process mining has been applied in the healthcare sector. However, upon reviewing related literature, it is evident that there are certain gaps to note and we have generalized those gaps into four. The primary gap identified is the predominant focus on case studies about process discovery techniques, particularly emphasizing control flow perspectives. Although the control flow perspective is the starting point, if we do not use it to analyze other perspectives such as performance and organizational perspectives, it may hinder comprehensive analysis and improvement of processes.

The second limitation is that the majority of research has focused on structured data, which may not accurately represent the dynamic and complex nature of healthcare environments. Structured data often fails to capture the nuances of patient-provider interactions, evolving conditions, and qualitative factors like patient experiences. Much valuable information exists in unstructured formats, which, if overlooked, can lead to incomplete insights and hinder innovation. To address this issue, it is essential to employ diverse preprocessing techniques such as aggregation, temporal approach and simple heuristic filtering and apply advanced process mining algorithms such as heuristic mining, inductive mining and social network analysis, capable of handling unstructured data effectively. Through these methods, unstructured healthcare data can be converted into structured formats, enabling a more comprehensive analysis of critical business processes within the healthcare sector.

The last gap that motivated to conduct this research is the dynamic [3] and context-specific nature [26] [27] [28] of healthcare environments. Researchers have suggested that different hospitals should consider applying various process mining techniques and methodologies. Given Ethiopia's unique context, we aim to demonstrate the potential for obtaining valuable insights into hospital management by implementing these methods in Ethiopian hospitals.

CHAPTER THREE

METHODOLOGY OF THE STUDY

3.1. Research Design

This research follows an experimental research approach to answer the research questions and to address both the general objectives and specific objectives of the study. Experimental research is a scientific approach to study, where one or more independent variables are manipulated and applied to one or more dependent variables to measure their effect on the latter [30]. The research aims to apply various process mining techniques to discover the healthcare process model. Each process mining techniques have its own strengths and weaknesses. Through experiment, the researcher has selected the best process mining techniques to discover the model that indicates detailed descriptions of actual healthcare processes. The experimental approach is not only enabled to select the best process mining techniques, but also to know how can we obtain insight in the healthcare process to identify bottlenecks and inefficiencies, improve patient care, and optimize resource allocation.

The study tries to identify significant attributes, from the event log, by applying different process mining techniques to identify the cause to the result of inefficiency and bottlenecks in the healthcare process through experiments.

3.2. Research Process Model

To conduct an extensive experiment, the researcher adopted the work of [31] with some modifications. From different process mining methodologies, this model is more appropriate for this research to analyze the healthcare process from control flow, performance, and organizational perspectives. This process model (see Figure 3.1) comprises data extraction, data preprocessing, process discovery, evaluating models' quality, control flow analysis, performance analysis, organizational analysis, and transfer results.

3.3. Data Extraction

The initial stage in data preprocessing for process mining involves data extraction. This process entails constructing the event log from the raw dataset by identifying relevant attributes to include, determining the timeframe for extracting event logs and selecting events with clear visibility [41].

The data was acquired from the TASH health information systems, known as IWKET and MedWeb. To create the single event log, the data is collected from both databases as csv file and merged them. The event data contains all the necessary information to visualize the actual process taking place in the hematological healthcare process. This data was obtained after receiving approval from the ethics board (see Annex A), ensuring that all procedures were conducted in a manner consistent with ethical standards.

3.4. Data Preprocessing

Once event logs are extracted from the data and integrated, the subsequent step involves filtering them. Event logs often present several issues such as irrelevant, incomplete, noisy, outlier, and missing data. Data preprocessing to address these issues is imperative to prepare the event log for further analysis [41]. Before applying process mining techniques to discover the model, the event data has been cleaned, filtered, and transformed into a format suitable for analysis. The study followed aggregation approach to handle and make a batched event as a single event to simplifies the process model and improves process mining quality. Another preprocessing approach the study followed is temporal approach, to improve the temporal aspect of the events. In this study, there are repeated activities that contributes to a complex process model that is difficult to understand. A simple heuristics technique was applied for data preprocessing, as it automatically selects events that frequently occurred in the event log. This technique is available in Open-source tools the latest ProM version 6.13, which is the most commonly used tool in healthcare studies [13].

3.5. Process Discovery

In this study direct implementation strategy has been utilized, which consists of applying process mining to a set of data gathered directly from HIS sources for building an event log [1]. Using this strategy consequently obtains models, tables, and diagrams for analyzing from different perspectives such as control flow, performance, and organizational perspectives.

The main techniques or algorithms used in the previous case studies to discover the healthcare processes are heuristics miner [32] [33] [34] and inductive miner [35] [36]. For the control flow model, heuristics mining and inductive mining are applied in this study. For performance analysis, using an inductive visual miner, we have discovered a model that shows the bottlenecks and inefficiencies that have to be improved to be more effective and efficient. For organizational analysis, a social network miner has been applied.

3.6. Process Analysis

For this study, the basic analysis strategy is employed. This is using process mining techniques and algorithms that are common in the available tools to an event log, without introducing any new techniques or algorithms [1]. The reason to use this strategy is that it can be the easiest to perform the analysis and less resource-consuming. This study conducts an analysis of control flow, performance, and organizational aspects of the process.

In this study, we have analyzed control flow of the healthcare processes to identify bottlenecks, and variations in care processes. It also enables the identification of deviations from standard protocols or best practices. Additionally, Performance analysis has been conducted by measuring key performance indicators (KPIs) such as patient waiting times, treatment durations, resource utilization, to identify opportunities for improvement, streamline processes, and enhance patient outcomes. Another perspective in this study is organizational analysis that helps to reveal insights into workflow dynamics, interdepartmental coordination.

3.7. Evaluating Models Quality

Measuring the quality of the discovered model is very important task for demonstrating the performance of algorithms applied to the event log. It helps to measure how well a process model describes the observed behavior [40]. However, one of the challenges in process mining involves balancing quality criteria such as fitness, simplicity, precision, and generalization [8]. Fitness refers to how well the discovered model permits on behalf of the activities noticed in the data set. Precision refers to the ability of a model to not permit the use of activities, which was entirely dissimilar in the data set. Generalization refers to how well a model performs on new and unseen data. A model that generalizes well will make accurate predictions on data it has not been trained on. Lastly, simplicity means the discovered model should be promising with the plain model.

In this study, process model quality metrics such as fitness, precision, generalization, and simplicity are evaluated for models discovered by the algorithms such as heuristic miner and inductive miner. The fitness metric is assessed using the ProM plugin called "Replay a log on Petri net for conformance analysis". Once the event log has been filtered, this plugin uses the petri net that the two discovery miners produced. The outputs are a "Result Replay" and “log contains statistical information such as trace fitness”. Precision and generalization metrics are measured using the "Measure the Precision/Generalization" plugin, which requires three elements (the Petri net, event log, and the result of fitness). It is acknowledged that there is no single perfect metric for quantifying simplicity [41]. Therefore, this study adopts the approach proposed in [41], measuring the metrics of density, cardoso, cyclomatic, and structuredness.

3.8. Transfer Result

In this stage, the findings from the previous phase have been interpreted to potential action points and evaluated by subject matter experts. The results of the healthcare process mining experimentation highlight the complexity and dynamic nature of healthcare services need for an integrated healthcare system that promotes continuous improvement through process mining techniques. Building on this insight, a healthcare improvement framework is proposed. This framework is then presented in detail to the participants involved in its evaluation. The evaluation process includes department heads and senior health and IT professionals, who were carefully chosen through purposive sampling. These experts rigorously evaluated the healthcare service improvement framework to ensure its reliability and relevance within the healthcare context.

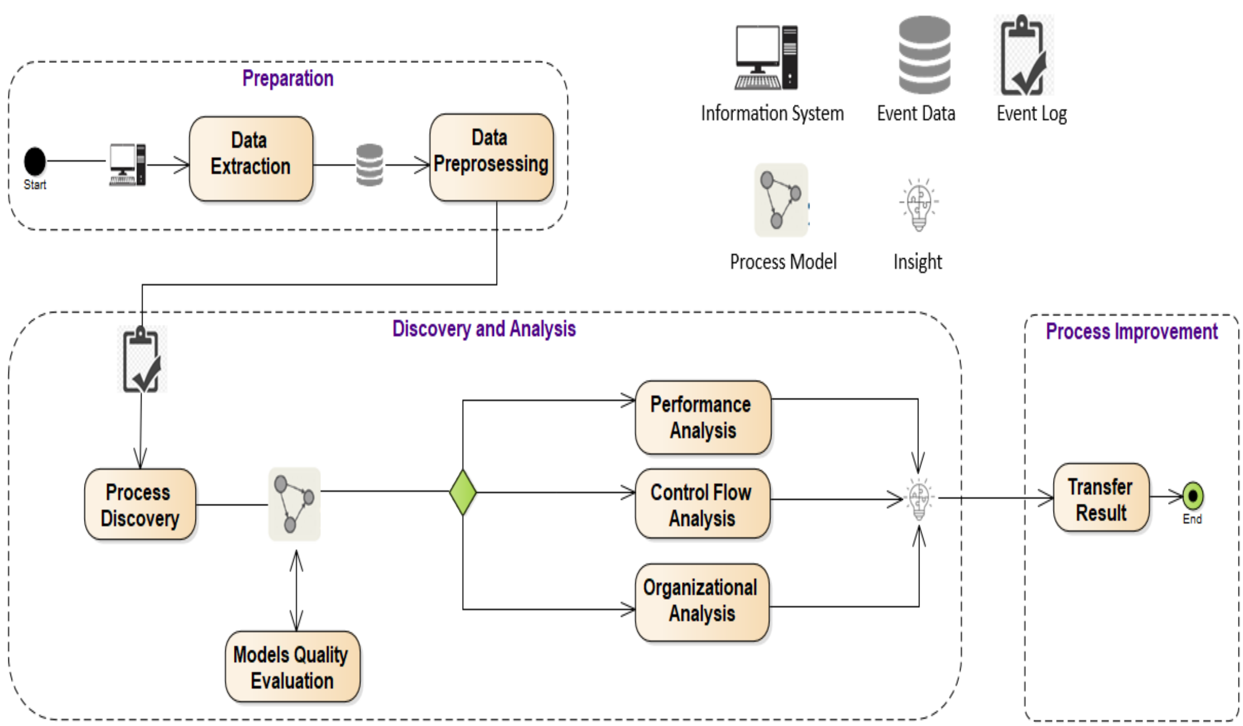


Figure 3.1: The proposed process mining model

CHAPTER FOUR

DATA EXTRACTION AND PREPARATION

4.1. Overview of the Business Process

Tikur Anbessa Specialized Hospital (TASH), also known as Black Lion Hospital, the main teaching and specialized healthcare facility in Addis Ababa, Ethiopia. Established in 1972, it offers unique clinical services not found in other Ethiopian hospitals [15]. TASH is staffed by 200 doctors, 379 nurses, 115 other health professionals, and 950 administrative personnel, with 800 beds to serve patients [46]. TASH has various departments, including the Hematology unit, which provides a comprehensive diagnosis and treatment of patients who have disorders of the blood and bone marrow, including different types of cancer, such as lymphoma, acute and chronic leukemia, multiple myeloma and myelodysplastic syndromes. Patients in the Hematology Unit receive care from a multidisciplinary team, including hematologists, specialized nurses, expertise in cancer, radiotherapists, internal medicine experts, surgeons, radiology specialists, pharmaceutical professionals, and infectious disease specialists.

The TASH hematology department has not created a handmade process model to plan patient care from admission to the end of their healthcare journey. As the department head says, the complexity of hematology diseases, which primarily affect blood cells and complicate treatment, makes it difficult to predict the care pathways patients should follow. For example, blood cancer impacts white blood cells, leaving patients vulnerable to other diseases and requiring referrals to other departments for further diagnosis and treatment. Despite this complexity, all hospital events are currently recorded in an information system. These event logs can be used to discover the actual healthcare process model using various process model discovery algorithms and techniques.

4.2. Ethical Considerations

Before collecting the data, ethical clearance is obtained from the Institutional Review Board (IRB) of Addis Ababa University, College of Health Sciences. The research protocol and data collection procedures were reviewed and approved by the IRB (see Annex B). We carefully removed irrelevant attributes like patient names and performer names that could potentially reveal personal identities. Additionally, the medical record numbers are anonymized to safeguard the privacy and confidentiality of the individuals involved in the study, ensuring compliance with ethical and legal regulations related to data protection and patient privacy.

4.3. Data Extraction

TASH uses two separate hospital information systems, such as IWKET and Medweb Image Server information system, to manage its business and data. The data is gathered from these two separated databases to create event logs for this study. The primary system, IWKET, stores most of the patient's information and journey. This system has been internally developed and currently supports outpatient services only; however, efforts are underway to expand its functionality to encompass inpatient services. Medical staff like doctors, nurses, laboratory technicians, cashiers, radiologists, and other professionals can input patient details, request services, record test outcomes and dispense medications through this system. All these activities are documented in a Microsoft SQL database. The data related to service request, diagnostics, and prescription are extracted from the database. Service request data includes events associated with physicians requesting laboratory tests and imaging/radiology services, while the diagnostics indicates the specific examinations and treatments received by patients.

The second system is referred to as Medweb Image Server, it is part of a robust imaging solution offered by Medweb that caters to the evolving needs of healthcare organizations worldwide. TASH uses this web-based system and stores data related to images such as X-rays, CT scans, ultrasounds, MRIs. The data is collected between 2020 to the April 2024, chosen for the significant amount of available information during this period.

In the following section, an explanation of the five event categories is provided along with a brief description of the sourced tables.

- **Diagnostic events:** The data was extracted from the IWKET database, specifically from the diagnosis table. This table includes the Medical Record Number (**MRN**), treatment unit (**Unit**), and detailed diagnostic activities, as well as the date and time of treatment (**Date/Time**). It also contains information about the doctor who made the diagnosis (**User Name**) and the role of the doctor (**Role**).
- **Request events:** A doctor request specific services like laboratory tests or radiology procedures to further investigate the patient's condition. This event is recorded in the database IWKET, specifically in the request table. The request table in IWKET databases includes

important attributes such as Medical Record Number (**MRN**), Specifies whether it is a laboratory test or a radiology test (**Title**), service request date and time (**Date**).

- **Laboratory test events:** In the IWKET database the laboratory table is a comprehensive repository, which contains the medical record number (**MRN**), the type of laboratory test (**Title**), and the date & time of the result (**Date/Time**).
- **Radiology test events:** This imaging data is stored on the Medweb Image Server. The extracted data attributes include Medical Record Number (**MRN**), type of image taken (**Modality** such as DX, CT scan, US, CR, MG & MR), date/time of image capture (**Date/Time**), and the radiologist identification number (**Performer**).
- **Prescription events:** This event is extracted from the prescription table in the IWKET database, which holds crucial information related to prescribed medications for patients such as medical record number (**MRN**), information about the prescribed medication (**Item**), timelines (**Date/Time**). The following Table 4.1 provides summary of the key attributes taken from five event categories of IWKET and MedWeb databases for the purpose of process mining.

4.4. Data Integration

Most of the process mining aspects addressed in this study require the consolidation of all event logs into a single unified event log. The data integration is centered around patients, as each table shares the same ID for a specific patient, making it straightforward to create a unified database.

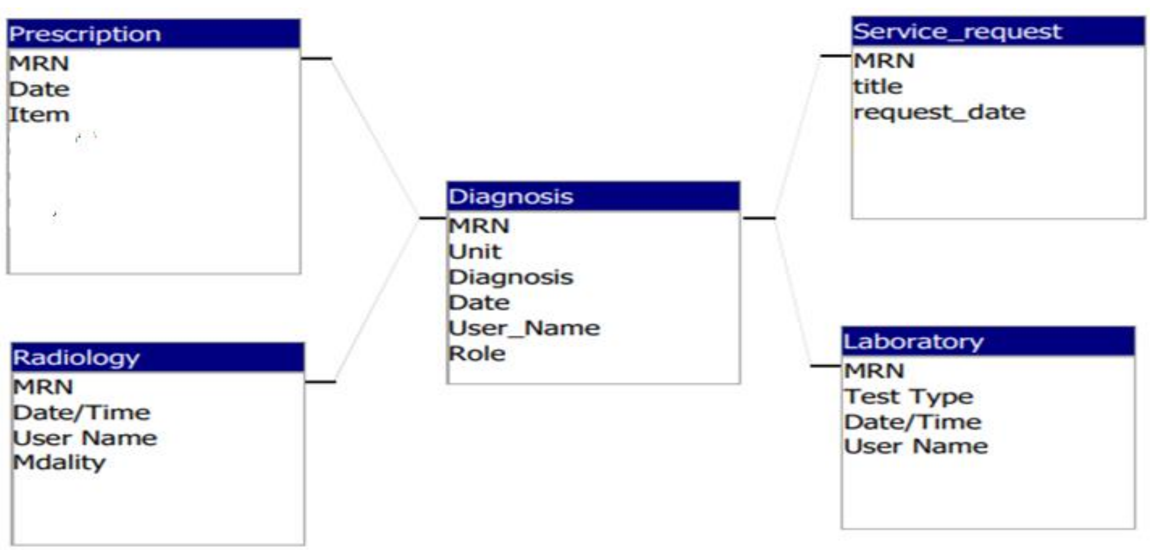


Figure 4.1: Data structures of the two information systems

Figure 4.1 shows the relationship between the tables that are extracted from the two information systems databases.

Based on the attributes, the data are combined into one single event log for our study as follows:

- **For Control Flow Analysis:** As discussed in Chapter Two, in order to identify the deviation and potential process optimizations of hematology treatment, the control flow perspective is analyzed by combining the diagnosis, request, laboratory, radiology, and prescription events are into a single event log. They share attributes such as MRN, Activities, and Date/Time for control flow analysis, as shown in Table 4.1.
- **For Performance Analysis:** The evaluation of the effectiveness, efficiency, and key performance indicators of hematology treatment process is performance analysis. to do that, the same as control flow analysis, the three common event log attributes (activity, timestamp, and case ID) are also utilized.
- **For Organizational Analysis:** To gain insights about the collaboration between departments, roles, and persons in the hospital, a social network analysis is conducted. To perform a department-level analysis, units were extracted from diagnosis events and other common departments including laboratory, radiology, and prescription. For role-level analysis, the role attribute is extracted from the diagnosis table; however, this attribute is absent in other event tables. Experts have suggested that laboratory events are typically carried out by laboratory specialists, prescription events by pharmacists, and radiology events by radiologists. Therefore, we labeled these event logs for role-level analysis. For employee-level analysis, only diagnosis and radiology event tables are used as they contain the name of the performer.

Table 4.1: Summary of key attributes considered for process mining

Table	Has Case_Id	Has timestamp		Has duration	Has activity name	Has Performer	
		Date	Time			ID	Role
Diagnosis	Yes	Yes	Yes	No	Yes	Yes	Yes
Request	Yes	Yes	Yes	No	Yes	No	No
Laboratory	Yes	Yes	Yes	No	Yes	No	Yes
Radiology	Yes	Yes	Yes	No	Yes	Yes	Yes
Prescription	Yes	Yes	Yes	No	Yes	No	Yes

4.5. Data Preprocessing

4.5.1. Filtering Hematology Patients

When the dataset was extracted, most event tables include all types of patients' data. Thus, a key task during preprocessing was to select only those patients whose treatment fell under the hematology unit. The criteria for this involved identifying patients who had been treated and monitored at least once in the hematology department, regardless of any other medical care they may have received. To pinpoint these individuals, from the diagnosis table basic filtering methods were utilized in Excel to locate those who had undergone hematology treatment. Following this identification process using the medical record numbers of these hematology patients, relevant event data was then extracted from other events tables such as prescriptions, requests, and radiology using advanced filtering methods in Excel.

Before preprocessing, there were a total of 5,579 instances identified as hematology patients with an overall count reaching 1,333,218 across various events belonging to 834 different classes. As discussed in the following section, applying different preprocessing approaches such as aggregation, temporal approach and simple heuristic filtering significantly reduced the process complexity. Unless applying preprocessing techniques, as shown in Figure 4.2, the discovered process model is considered a "spaghetti-like" or unstructured model that seems hard to understand to conduct further analysis.

Figure 4.2: The discovered process (a spaghetti-like) model before preprocessing

4.5.2. Data Aggregation

In the database, prescription and laboratory events contain a large number of batch events which should be addressed as a preliminary step for mining patient pathways. Batched events are re-extracted with the same event label, as shown in Table 4.2. For process mining purposes, in consultation with experts, events such as CBC, ESR, MORPH, and others (like INR, PT, CCD, APTT etc.) are labeled under the name of Laboratory event. Another events log such as CR, DX, CT, and others (like MG, MR, US etc.) under the activity name radiology. The event log prescription also follows the same fashion. This aggregate preprocessing approach has notably reduced the large number of event classes from 834 to 38, as shown in Table 4.4.

Table 4.2: Mapping event name for simplicity

Events name	**Mapped event name**
CBC	Laboratory test
ESR	Laboratory test
MORPH	Laboratory test
Others (like INR, PT, CCD, APTT etc.)	Laboratory test
CR	Radiology
DX	Radiology
CT	Radiology
others (like MG, MR, US etc.)	Radiology
Allopurinol	Prescription
Aceclofenac	Prescription
Rituximab	Prescription
Others (like Meropenem, Syringe, Cefazoline etc.)	Prescription

4.5.3. Handling Repeated Temporal Events

This method aims to improve the temporal aspect of the events. In this study, there are repeated activities in the case of prescription and laboratory events. As shown in Table 4.3, the prescription event log includes ordered for different items for the specific patient at the same time, recorded as separate activities, contributing to a complex process model that is difficult to understand. Similarly, the laboratory event log faces similar issues by listing different laboratory tests in the event log at the same time. To address these complications, all items prescribed to a patient at the same time are considered as one event.

Table 4.3: A list of items prescribed to patient at the same time

MRN	Items	Date	Items (after aggregation)	Date (after temporal approach)
12346060	Syringe (Disposable) - 10.00 milliliter	7/9/2023	Prescription	7/9/2023
12346060	Sodium Chloride - 1000ml	7/9/2023		
12346060	Meropenem - 1,000.00 Milligram	7/9/2023		
12346060	Meropenem - 1,000.00 Milligram	7/9/2023		
12346060	Meropenem - 1,000.00 Milligram	7/9/2023		
12346060	Adrenaline (Epinephrine) - 0.10 % in 1ml	8/1/2023	Prescription	8/1/2023
12346060	Cefazoline - 1.00 Gram	8/1/2023		
12346060	Syringe (Disposable) - 50.00 millilitre	8/1/2023		
12346060	Syringe (Disposable) - 5.00 millilitre	8/1/2023		

The total number of events decreased from 1,333,218 to 258,832 through temporal preprocessing, resulting in a substantial impact on the process model. Additionally, the mean events per case and maximum events per case were also reduced by using this method. The use of these different approaches resulted in noteworthy changes across various characteristics within the event log data. This highlights how effective they are at streamlining the process model and reducing overall complexity.

Table 4.4: Event log characteristics and preprocessing steps

Event log characteristics	Raw log	Data Aggregation	Handling Temporal Events
Patients (Cases)	5579	5579	5579
Event classes	834	38	38
Events	1,333,218	1,333,218	258,832
Mean event per case	239	239	46
Minimum event per case	1	1	1
Maximum event per case	1862	1862	472

Table 4.4 shows that the summery to show an aggregate preprocessing approach and temporal preprocessing approach have led to significant changes in all event log characteristics such as number of events, event classes, and the number of events per case.

4.5.4. Filtering Incomplete Process Instances

Despite the reduction in process model complexity from the initial pre-processing, achieving a well-structured process model requires advanced pre-processing and individual-level trace filtering. The study uses a simple heuristic to remove those incomplete process traces from the log and eliminate all traces that do not begin and/or end with a particular event. Table 4.5 shows the event log data statistics before and after applying this heuristic filter.

Table 4.5: Event log before and after applying simple heuristic filtering

Filtering status	No. of cases	No. of events	No. of event classes	No. of mean events per cases
Before apply filter	5579	258,832	38	46
After apply filter	3804	177700	4	47

As shown in Table 4.5, applying the default settings of the simple heuristic miner filtered in 80% of the data, leading to a reduction in the number of cases from 5,579 to 3,804. Furthermore, the number of events decreased from 258,832 to 177,700, and the number of event classes is significantly reduced from 38 to 4.

The following lines of a comma-separated value (CSV) file is a sample dataset used in the experimentation. It includes key attributes such as medical record number (MRN), activity, timestamp, along with the respective values for each attribute.

```
MRN,Activity,Timestamp
12346060,Laboratory Request,7/6/2022 0:00
12346060,Hematology,8/31/2022 0:00
12346060,prescription,8/31/2022 0:00
12346060,Laboratory Request,8/31/2022 0:00
12346060,Laboratory Test,8/31/2022 9:52
12346060,Hematology,9/28/2022 0:00
12346060,Laboratory Request,9/28/2022 0:00
12346060,Laboratory Test,9/28/2022 10:29
12346060,Hematology,10/26/2022 0:00
12346060,Laboratory Request,10/26/2022 0:00
12346060,Laboratory Test,10/26/2022 8:20
12346060,Hematology,11/9/2022 0:00
12346060,prescription,12/29/2022 0:00
12346060,prescription,12/30/2022 0:00
12346060,Laboratory Request,12/30/2022 0:00
12346060,prescription,12/31/2022 0:00
```

CHAPTER FIVE

EXPERIMENTATION AND RESULT ANALYSIS

In this chapter, the focus is on understanding how to discover and analyze different process models within a given context. This chapter delves into the systematic approach of discovering existing process models based on data extracted from the information system within an organization. It emphasizes the importance of thorough exploration and examination to identify key elements, relationships, and dependencies within these models. By following structured methods and techniques such as heuristic miner and inductive miner, professionals gain insights into the underlying processes that drive organizational operations. The chapter also covers strategies for analyzing these models to improve efficiency, identify bottlenecks, and optimize workflows for better performance.

For the experimentation, we utilized process mining and process analytics techniques on the event logs generated from the data preprocessing stage. To apply process mining techniques, we employed ProM version 6, an extensible open-source framework [85]. The framework provides easy-to-use user interface functionality, a variety of model-type implementations (Petri nets, BPMN notations, YAWL), the loading and filtering of event logs, several process mining algorithms, and the corresponding visualization of results.

5.1. Data Set for The Experiment

For this study, the event log is extracted from separate information systems and integrated into one single dataset, and then the extracted dataset is preprocessed to select only hematology patients and to get a more structured and easily understandable process model. The input data for these algorithms are collected from the diagnosis, request, laboratory, radiology, and prescription events by making one single event log. The attributes of this extracted event log are case ID, activities, timestamp, department, role and performer.

The study utilizes data from 3,804 cases, 177,700 events, and 4 event classes to analyze the control flow and performance perspectives. Both heuristic and inductive miners are employed to leverage this data for understanding the sequence and performance of activities within the process. Additionally, for social network analysis, the study uses data comprising 4 departments, 5 roles, and 999 performers.

5.2. Process Analysis Using Heuristic Mining

One of the proposed process mining algorithms for this study is the heuristic miner plugin in ProM. The heuristic miner algorithm handles noisy and incomplete data by applying heuristics to reconstruct the process flow. The goal of using this algorithm is to create a model that best fits observed behavior in the log, making it valuable for analyzing control flow perspectives of hematology treatment processes.

The first step of the heuristic miner is creating a dependency graph as shown in Figure 5.1. To create a dependency graph, the dependency matrix and the length one loop dependency are built. It builds the directly follows matrix by computing the frequency between the activities using Equation (1). Figure 5.1 is dependency graph that represents the process model of hematology treatment at TASH. The heuristic miner has introduced an artificial start to represent the beginning of traces and end activities within the graph. The graph illustrates that some cases start with a laboratory test, others with a prescription, some with hematology, and others with a laboratory request. These four events are connected to the artificial start, which is depicted as a dark rectangle at the top of Figure 5.1. Similarly, the endpoint is linked to three activities: prescription, laboratory test, and laboratory request, it shows that some cases are finished their events after laboratory test, some others are finished with prescriptions and the remaining ended with laboratory request.

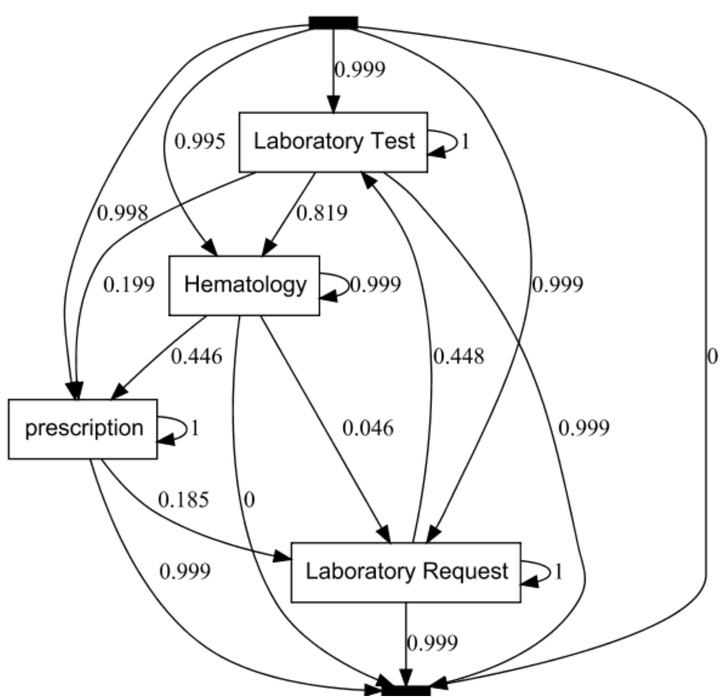


Figure 5.1: Results of heuristic miner: The dependency graph

The decimal values presented in Figure 5.1 indicate the dependencies between different events. Any value greater than 0 signifies a relationship between two events. For example, there is a relationship between a laboratory test and a prescription, represented by a value of 0.199, between a laboratory test and hematology with a value of 0.819, and between a laboratory request and a laboratory test with a value of 0.448. Values closer to 0 suggest that the relationship between two events is bidirectional, while values closer to 1 indicate a more unidirectional relationship. A value of 0 signifies that there is no relationship between the events. In Figure 5.1, the value between hematology and an artificial endpoint is 0, indicating that no cases conclude with the hematology event. One advantage of this algorithm is its ability to output a heuristic net that can be changed into other types of process models for further analysis in ProM [85]. Based on the dependency graph, heuristic miner generated the directly follow graph as shown Figure 5.2, that shows control flow perspective of the healthcare process.

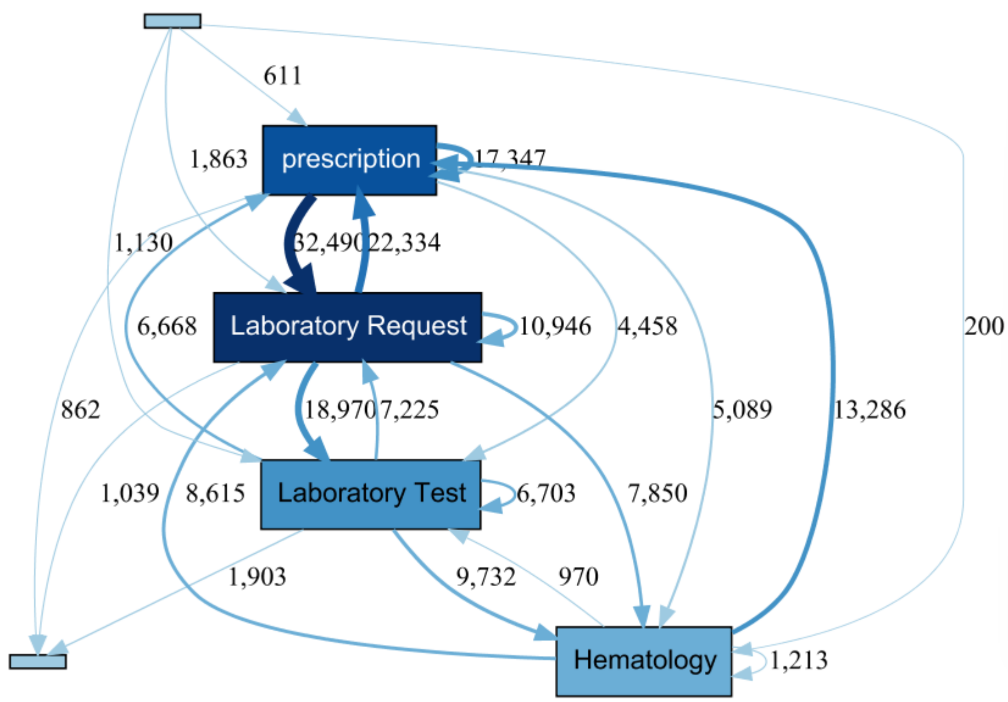


Figure 5.2: Results of heuristic miner: Directly follow graph

The process model shown in Figure 5.2 illustrates how the actual hospital processes take place. A total of 3,804 patients were observed in the model, with the majority, 1,863 patients, beginning their treatment with a laboratory request. This suggests that, in most cases, patients are required to undergo laboratory testing prior to diagnosis or other procedures, using the test results to guide subsequent treatment decisions. The activities are represented by boxes, while the transitions between activities are visualized with arrows. The numbers on the arrows indicate the frequency of patient transitions between these activities. Vivid colors rectangle shape highlights the activities with the highest frequency in the log; for example, the most frequent activities are laboratory requests, followed by prescriptions. The arrow with the greatest thickness and most vivid color corresponds to the most highly connected events. The most common transition occurs from prescription to laboratory request with 32,490 instances. This is predictable since after taking medication often comes further diagnosis through laboratory testing. Additionally, the flow of patients from one activity to another is ranked as follows: 22,334 patients from laboratory request to prescription, 18,970 patients from laboratory request to laboratory test, 17,347 patients from prescription to prescription, and 13,286 patients from hematology to prescription. Figure 5.2 also includes details of the graph, highlighting less frequent connections. The least common patient flow starts from hematology, with only 199 patients.

It is evident from the model that there can be variations in the sequence of events for different patients. The fact that a significant number of patients started and ended with different processes indicates a complex and possibly fragmented patient journey. Healthcare professionals have confirmed that typically patients undergo a laboratory request followed by laboratory test and its results being reviewed at the hematology department for disease diagnosis, leading to prescription medication or additional laboratory tests if necessary. The findings shown in Figure 5.2 are consistent with expert expectations. As illustrated in the figure, the most common pathway for hematology patients begins with a laboratory request, followed by a laboratory test, then a hematology diagnosis, and finally a prescription. In addition to this typical sequence, there are variations in the pathways some patients follow. For instance, some patients start with a laboratory test, proceed to hematology, and then receive a prescription before concluding their treatment.

5.3. Process Analysis Using Inductive Mining

Besides heuristic miner, the inductive miner is also experimented that able to return sound and fitting process models and reproduce all observed behavior in finite time (see Figure 5.3). By applying the inductive visual mining, information about the distribution of the patients among different departments is obtained as shown in Figure 5.3, which depicts the pathway that hematology patients follow for their treatment.

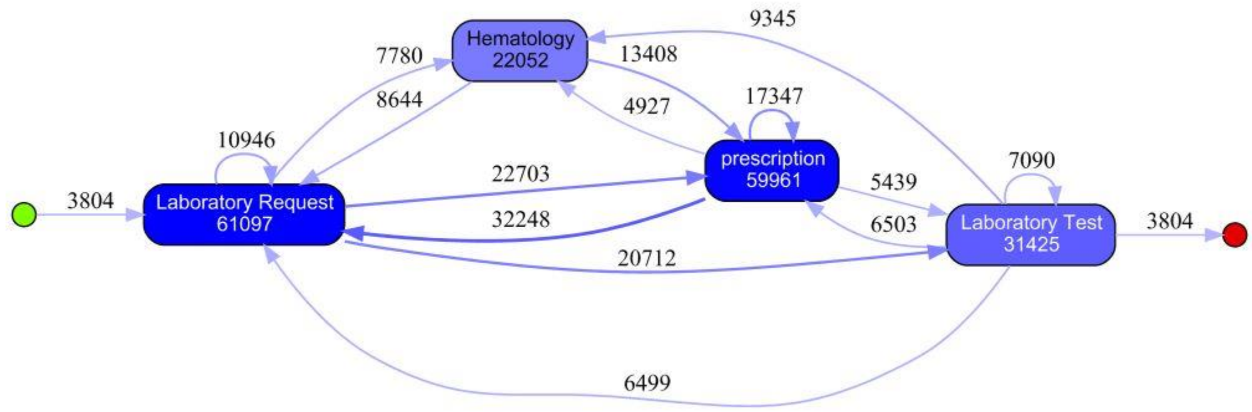


Figure 5.3: Patient distribution among different activities using inductive visual miner

Figure 5.3 depicts that 3,804 cases follow various paths. The data shows that most of the patients initiated their service with a Laboratory request, and most of them concluded it with a Laboratory test. Some patients repeat certain activities or recycle within a particular pathway. The most frequent activity observed is "laboratory request," occurring 61,097 times, suggesting that one patient may undergo multiple laboratory requests. Following closely behind with an occurrence of 59,961 times is "prescription," indicating that a patient frequently take medication. The frequency of laboratory requests and the actual completion of laboratory tests are expected to align closely. However, the findings reveal a notable discrepancy, suggesting that some patients do not undergo the laboratory tests as requested. Additionally, there are instances where laboratory results are not recorded in the system; instead, the laboratory personnel provide the results to the patients on paper. To address these issues, it is essential to ensure that all requested laboratory tests are completed and properly documented. Similar to the heuristic miner model described earlier, lab requests and prescriptions are also commonly interconnected with prescription to laboratory request interactions occurring 32,248 times while laboratory request to prescription occurring 22,703 times.

As depicted in Figure 5.4, the inductive visual miner demonstrates the movement of each patient from the initial point to their endpoint along with specific activity arrival times. The green dot signifies each patient's starting point as they move along their path until reaching the red endpoint. The yellow dots represent individual patients following distinct pathways within the model. Some activities show loops, indicating frequent repetition by patients: laboratory requests occur 10,946 times, prescriptions are repeated 17,347 times and laboratory tests are repeated 7,090 times. As previously stated, these activities are repetitive in the order prescriptions, laboratory requests and then laboratory tests.

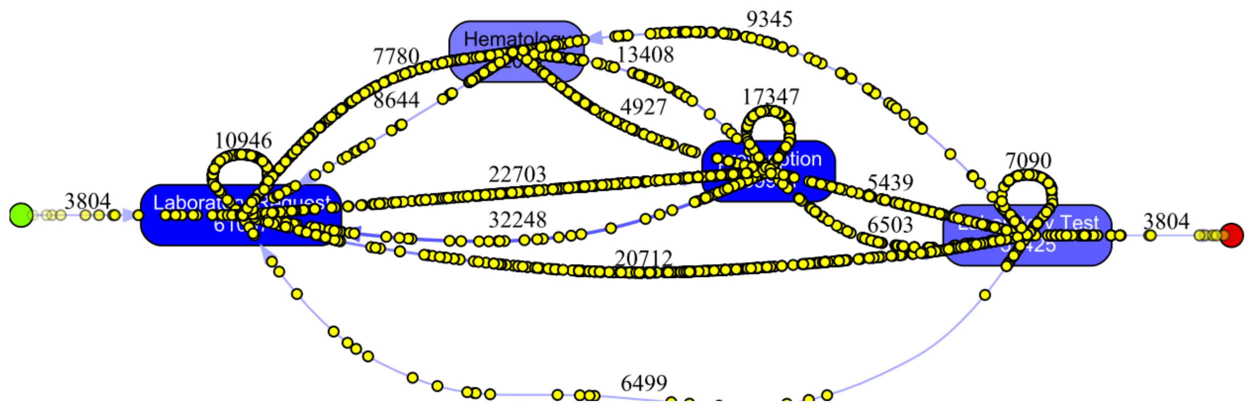


Figure 5.4: patient flow using inductive visual miner

5.3.1. Performance Analysis Using Inductive Mining

Performance analysis of hospital's process flow is very crucial to determine how much time the patient has spent since arriving at the hospital, how long each activity takes, how many resources are used, which patient activities take up a lot of time, and how long the patient must wait for a specific procedure.

In order to examine the healthcare processes from a performance perspective, we have utilized inductive visual miner. Figure 5.5 depicts the patients who received hospital services during the observed time frame. There is a high volume of patient flow from prescription to laboratory request, occurring 32,248 times. This suggests that when these units are in close proximity, it enhances the ease and speed of access for patients, thus reducing their treatment duration. Notably, Laboratory request emerged as the most frequently executed activity at 61097 instances, indicating high workload within this task. Consequently, resource allocation should be carefully considered for this particular activity. Moreover, Prescription stood out as the most frequently repeated activity with 17,347 occurrences. This indicates multiple batches of medication are being provided to each patient at once.

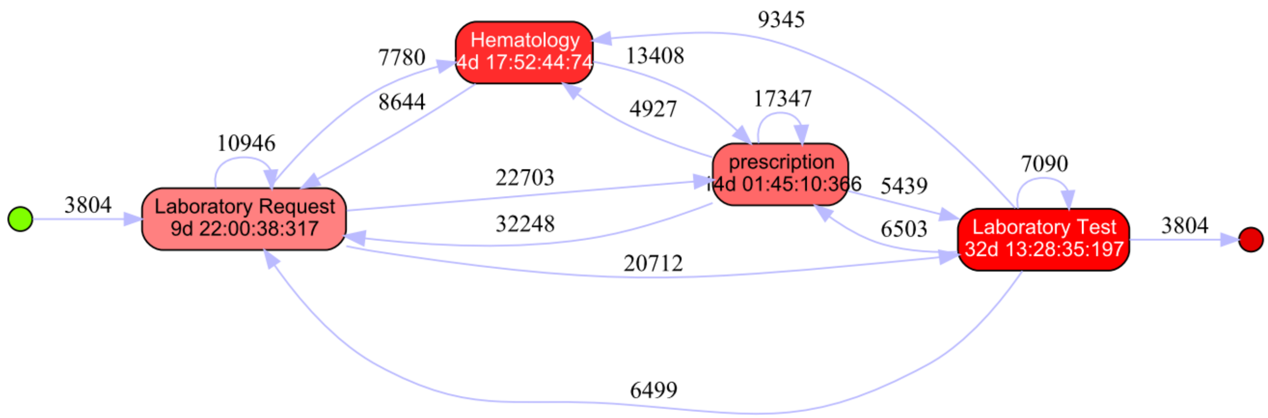


Figure 5.5: Hematology patients' path way and sojourn time

There is also a way in the inductive visual miner to observe where time is spent mostly in different ways. One of them is to visualize the sojourn time of activities. The activities denote how long execution took on average. As shown in Figure 5.5, we see that the longest time taken activity is the Laboratory test which takes on average 32 days followed by hematology on average 24 days and Serology with an average of 20 days as depicted below it. Based on the findings, it should be carefully investigated why it takes a long time for execution. World Health Organization (WHO) emphasizes the importance of timely diagnostic services, including laboratory tests, as a critical component of quality healthcare. For non-emergency cases, WHO suggests that diagnostic tests should be completed within a few days, and certainly not extending into weeks. The prolonged sojourn times observed in this study (e.g., 32 days for laboratory tests) would significantly exceed typical benchmarks, indicating a serious bottleneck [69]. The hospital management has acknowledged that the delays in laboratory test is primarily due to an imbalance in resource allocation. Patients are required to wait for extended periods until their scheduled appointments day. Typically, they undergo their laboratory tests one day before their appointment and then proceed to the hematology department with their results. This process creates bottlenecks in healthcare services. Therefore, it is imperative for the management to address resource allocation in the hematology department by hiring additional doctors and nurses to expedite patient care and reduce waiting times.

5.4. Evaluating the Models

This study compared among heuristic miner and inductive miner to select one with the high scores in the quality metrics (fitness, precision, generalization, simplicity). We evaluate the quality of the process models generated in this study, which are the resulting models obtained from applying the heuristic miner and inductive miner discovery algorithms.

5.4.1. Fitness Metric

The discovered process model's fitness is measured using the alignment-based conformance checking method, used the ProM plugin named "Replay a log on petri net for conformance analysis" that was derived from the alignment-based conformance checking in [67]. This plugin takes the petri net that generated from the two discovery miners and the event log after filtering, the outputs from this plugin are a "Replay result" and "log contains statistical information such as trace fitness". The Figure 5.6 and Figure 5.7 shows the result of replaying the petri net from applied discovery miners with extracted log after filtering.

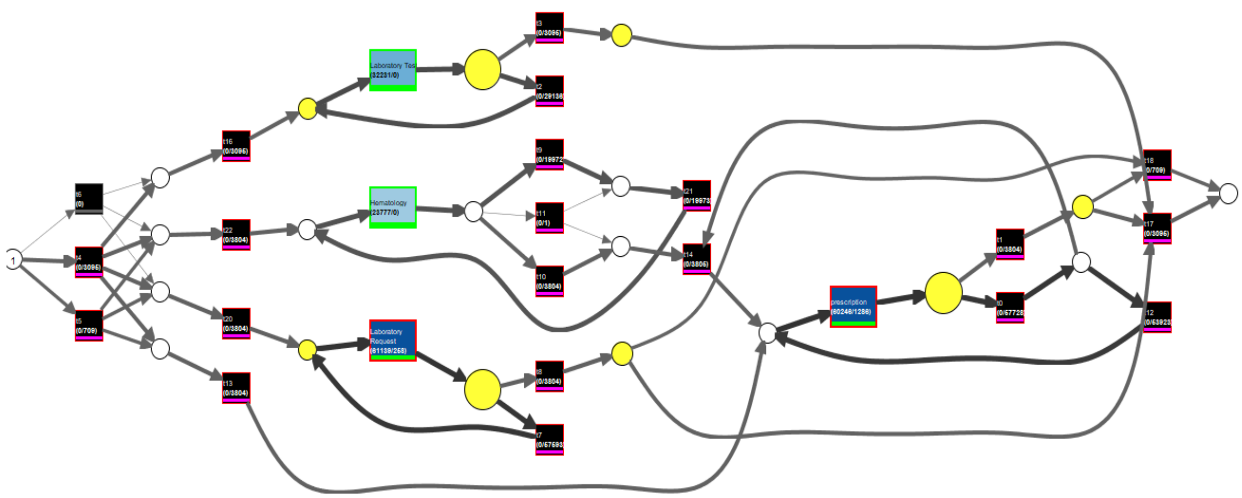


Figure 5.6: Replay result of Petri net based on heuristic miner with event log

The results of the heuristic miner, as illustrated in Figure 5.6, depict the process flow with the circle labeled "1" representing the starting point, and the final circle indicating the process endpoint. The blue rectangular shapes correspond to individual activities within the process. For the activities "laboratory test" and "hematology," there is complete alignment during the replay, as indicated by the green border. However, the "laboratory request" activity exhibits misalignment in 268 out of 61,138 activities. Similarly, the "prescription" activity shows misalignment in 1,288 out of 60,248 activities. Additionally, the yellow circle, white circle, and black rectangle shapes in the figure represent joins, and split of the transitions.

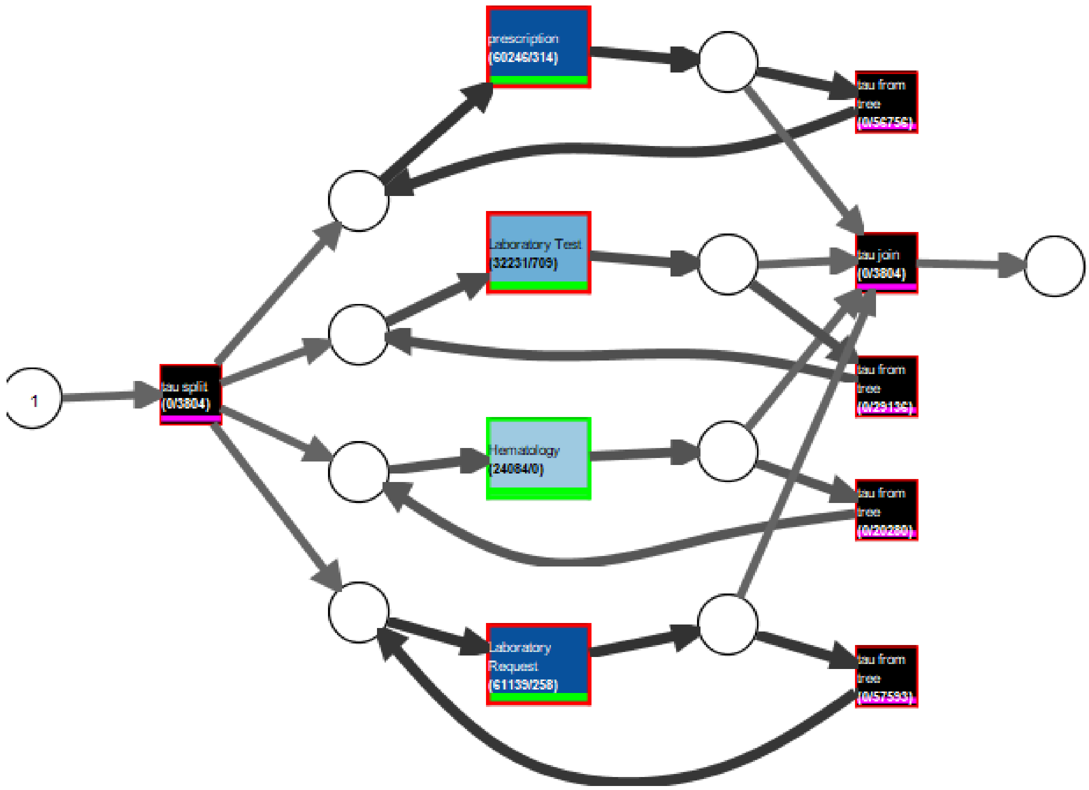
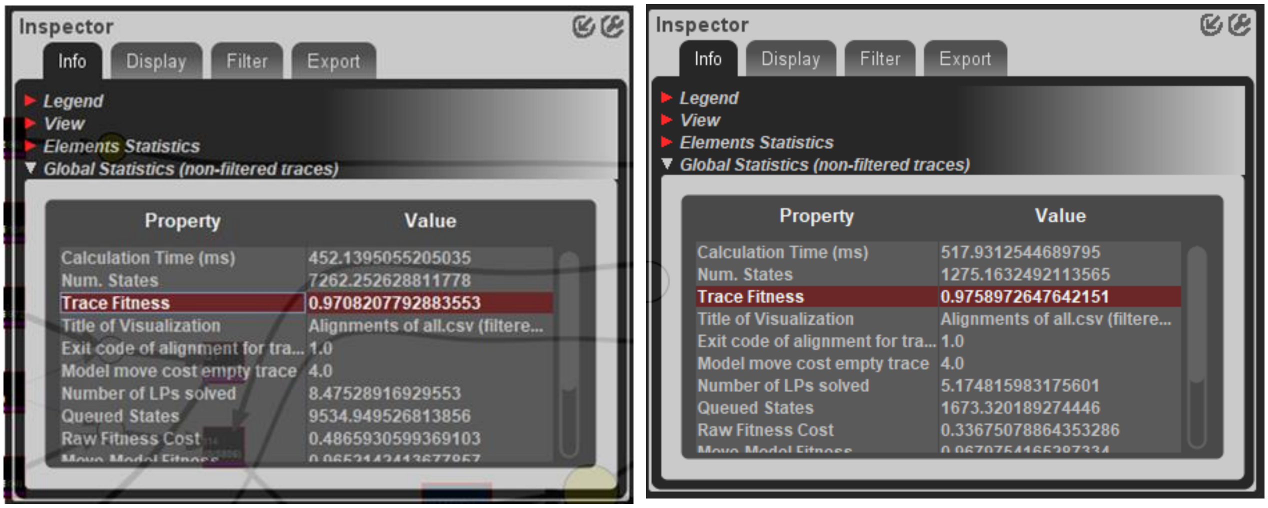


Figure 5.7: Replay result of Petri net based on inductive miner with event log

The output of the inductive miner replay, as depicted in Figure 5.7, indicates that the "hematology" activity exhibits no misalignments during the replay. However, all other activities display varying degrees of misalignment. Specifically, the "prescription" activity has 314 misaligned instances out of 60,246, the "laboratory test" activity has 709 misaligned instances out of 32,231, and the "laboratory request" activity shows 258 misaligned instances out of 61,139.

Another outputs in Figure 5.8(a) and Figure 5.8(b) shows the statistical information from replay petri net with the extracted log. The statistical data presented in the Figure 5.8 (a) demonstrates that the trace fitness of the heuristic miner is 0.9708, indicating that 97.08% of the traces in the event log align with the model produced by this mining technique. Similarly, in Figure 5.8 (b), the inductive miner also shows a trace fitness of 0.9759, signifying that 97.59% of the traces in the event log are consistent with the model generated by the inductive miner. This means that the inductive miner better aligns with the observed data or the event logs.



(a) (b)

Figure 5.8: Statistical information from replay Petri net based on (a) heuristic miner & (b) inductive miner with event log

5.4.2. Precision and Generalization Metrics

Three components, such as the petri net, event log, and result replay are required as input to measure the precision and generalization metrics of heuristic miner (see Figure 5.6) and inductive miner (see Figure 5.7).

The plugin of "Measure the Precision/Generalization" measures the precision using Equation (8), and generalization using Equation (6) that are introduced in [68]. Table 5.1 display results from the "measure precision/generalization" plugin for a petri net based on heuristic miner and inductive miner, respectively.

Table 5.1: Precision and generalization results of Petri net for heuristic and inductive miner

Algorithm	Precision	Generalization
heuristic miner	0.43736	0.99929
inductive miner	0.47878	0.99561

Upon examining the results as shown in Table 5.1, it is evident that the inductive miner outperforms the heuristic miner marginally in precision with score of 0.47878, as compared to the heuristic miner's score of 0.43736. On the other hand, the generalization score of 0.99929 registered by heuristic miner is slightly higher than the score 0.99561 of inductive miner, indicating both are a better ability to extract patterns applicable to new data. These findings suggest that while both algorithms demonstrate reasonable performance, the inductive miner exhibits a slightly superior capability in terms of precision and almost equal in adaptability to new datasets.

5.4.3. The Simplicity Metrics

This study uses a method proposed by [41], to quantify the simplicity metric and select the simple model from the heuristic miner and inductive miner. The method is summarized into two steps: Firstly; checking the soundness of every model resulted from the miner. Then second step; measuring the metrics of cardoso, cyclomatic, and structuredness.

Check the soundness

The study utilized the ProM plugin “Analyze with Woflan” to assess the soundness of two modes produced by heuristic miner and inductive miner algorithms shown in Figure 5.6 and Figure 5.7. A process model must meet specific criteria for a thorough soundness check, including being a workflow net. This involves fulfilling requirements such as having a single source place, a single sink place, and ensuring that each node is part of some path from the source place to the sink place.

In our initial experiment using the heuristic miner, the analysis revealed that the resulting model was considered "not sound" because it failed to meet one of the three soundness criteria: proper completion. The subsequent experiment focused on evaluating the inductive miner process model to determine its soundness. The result confirmed that "The net is a sound workflow net," as it satisfied all the criteria of a workflow net, as well as the requirements for a sound process model.

Measuring the metrics of density, Cardoso, cyclomatic, and structuredness

As we discussed in the research method part, the study used ProM plugin “Show Petri net metrics” to get the values of the simplicity metrics such as density, cardoso, cyclomatic, and structuredness. Table 5.2 shows the output values of the metrics for the models based on the heuristic and inductive discovery miners shown in Figure 5.6 and Figure 5.7. As shown in Table 5.2, The comparison between heuristic miner and inductive miner shows that the inductive miner, with a higher density value (0.13 vs. 0.062), produces more interconnected models, indicating complexity in relationships between elements. However, the inductive miner maintains overall lower complexity as indicated by lower cardoso (13 vs. 26) and cyclomatic (51 vs. 362) complexities compared to the heurist miner. Additionally, the heurist miner has a significantly higher structuredness value

(98150 vs. 350), suggesting its models are more complex and less organized. Thus, the heuristic miner generates more detailed but complex models, while inductive miner produces simpler, more structured, and manageable models.

Table 5.2: Quantify the complexity of the model

Model based on	Density	Cardoso	Cyclomatic	Structuredness
Heurist miner	0.062	26	362	98150
Inductive miner	0.13	13	51	350

Table 5.3 shows that the inductive miner slightly edges out the heurist miner in fitness (0.9759 vs. 0.9708) and precision (0.48 vs. 0.44), indicating it more closely aligns with the observed data and is better at minimizing false positives. However, the heurist miner has a marginally higher generalization score (0.99929 vs. 0.99561), suggesting it may perform slightly better with new or unseen data. A significant distinction is in their simplicity, the inductive miner is described as simple, which typically implies easier implementation and interpretation, whereas the heurist miner is considered complex. Overall, the inductive miner is the preferred choice because it offers a sound model that aligns closely with the data, ensures precision, and easy to use.

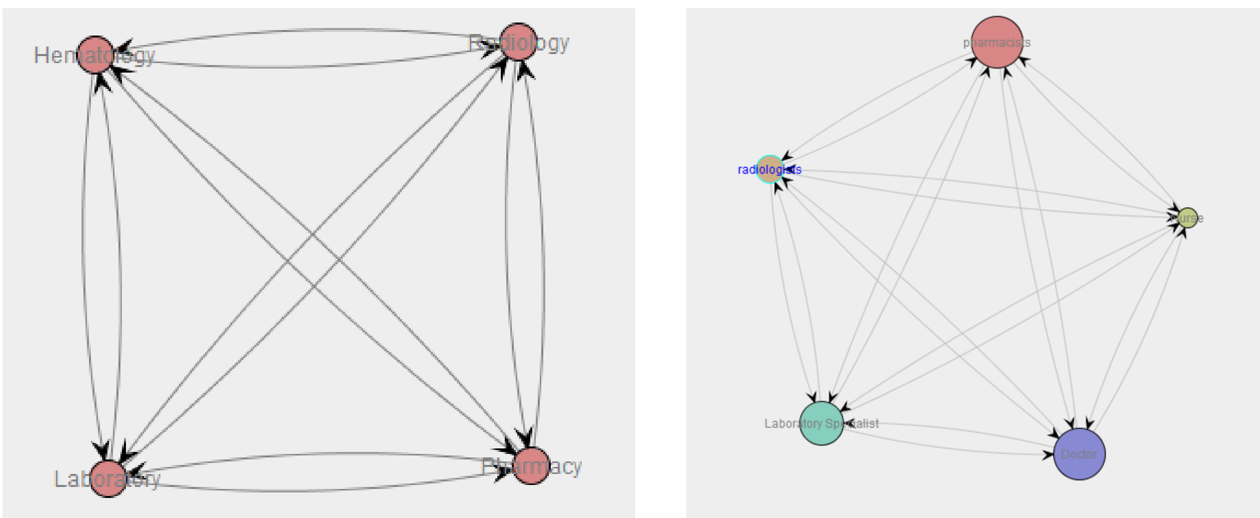
Table 5.3: Models evaluation summary between the two applied algorithms

Model based on	Fitness	Precision	Generalization	Simplicity
Heurist miner	0.9708	0.44	0.99929	Complex
Inductive miner	0.9759	0.48	0.99561	Simple

5.5. Social Network Analysis

Discovering social network helps to analyze organizational perspective of the healthcare process in three different levels such as the department level, the role level, and resource level. The technique discussed in this study mostly focus on the performer class which is called resource, department and role in addition to case id, activity, timestamp. The study used social network miner that plugged into ProM platform to gain insights about the collaboration between departments, roles, and persons in the hospital. Figure 5.9 shows the social network that generated from the event log, it used the “Handover of Work” metric that measures the frequency of transfers of work among departments. As shown in Figure 5.9(a), all departments such as laboratory, hematology, and prescription and radiology have strong interactions and frequently hand over work to each other.

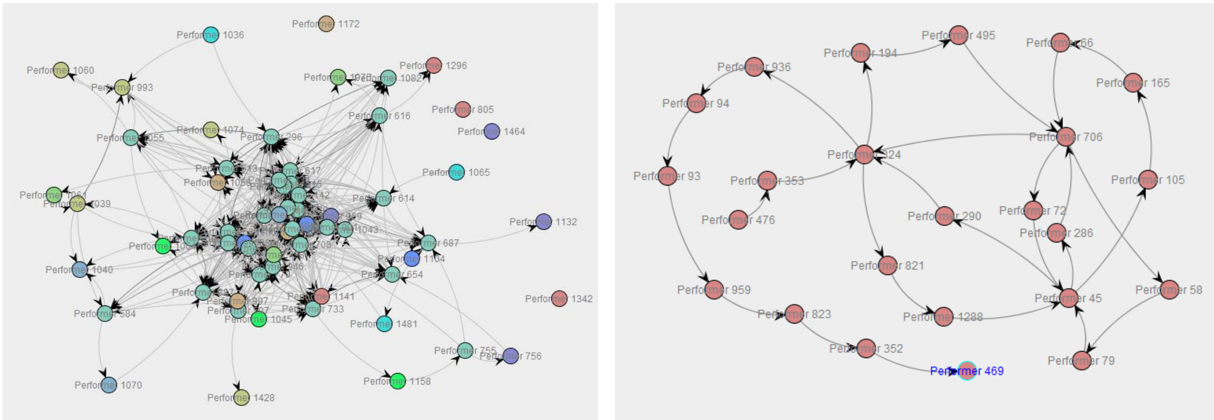
Figure 5.9(b) shows a complex network within a healthcare setting, with roles such as radiologist, nurses, doctors, pharmacists, and laboratory specialists interconnected. Radiologists are central to medical imaging and diagnostics, interacting with other roles like doctors, nurses, pharmacists, and laboratory specialists. Nurses coordinate patient treatments, administer medications, and facilitate communication. Doctors diagnose illnesses, prescribe treatments, and oversee patient care plans. Pharmacists manage medications and ensure patients receive appropriate ones. Laboratory specialists conduct medical tests, collaborating with doctors and nurses to order specific tests and communicate findings.



(a) (b)

Figure 5.9: Social network of handover of work among: (a) departments & (b) roles

The social network model, the larger circles represent events that occurred more frequently in the event log, while the smaller circles correspond to events that occurred less frequently. As shown in Figure 5.9(b), doctors and pharmacists are responsible for most of the events recorded in the log, whereas nurses and radiologists are associated with events that occur less frequently. We also discover the social network of handover of work among resources for a single patient and more than one patient, as shown in Figure 5.10.



(a) (b)

Figure 5.10: Social network of handover of work among resources for (a) many patients & (b) a single patient

Figure 5.10 depicts the transfer of tasks among healthcare professionals to illustrate the workflow relationship. Figure 5.10(a) illustrates that most resources, such as performers 296, 1104, and 1043, who are positioned at the center of the graph, frequently exchange work among one another. Conversely, certain resources, such as performers 1172, 1296, and 1345, neither hand over work to others nor receive it. Additionally, some resources, like performers 132 and 1428, receive work without passing it on, while others, such as performers 1060 and 1036, assign work without accepting any in return. Figure 5.10(b) illustrates the flow of work among resources involved in a single patient care journey. A total of 24 performer contribute to the overall process, with most resources being linked to only one other resource. However, some resources are connected to two or more others, such as performers 324, 206, and 45. The task is initiated by performer 476 and concluded by performer 469.

In our experimentation, another SNA plugin we applied was Similar-Task mining which creates connection between resources that frequently perform the same activity. We do this for radiology department as shown in Figure 5.11. As shown in Figure 5.11, the resources connected with similarity of activities performed by those resources. Connected components are assigned the same color, representing specific roles within the process. To identify the types of roles corresponding to these clusters, one can refer back to the event log to determine the activities performed by the resources within each cluster.

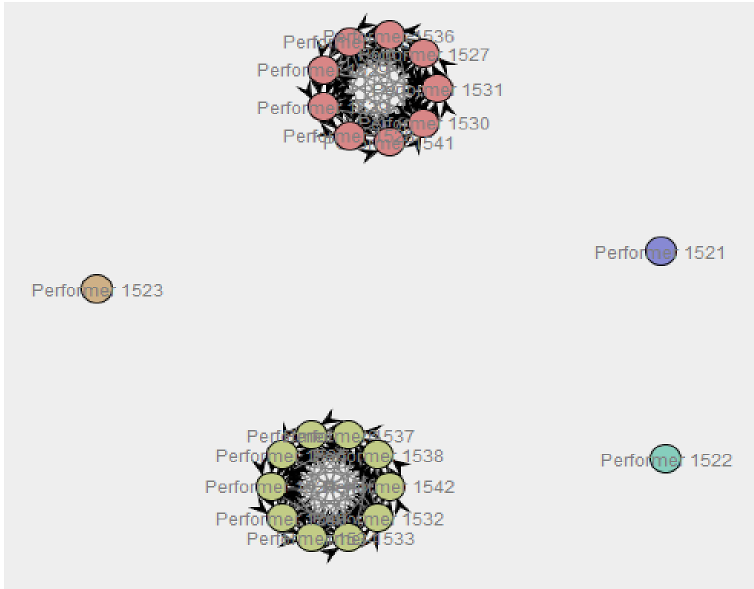
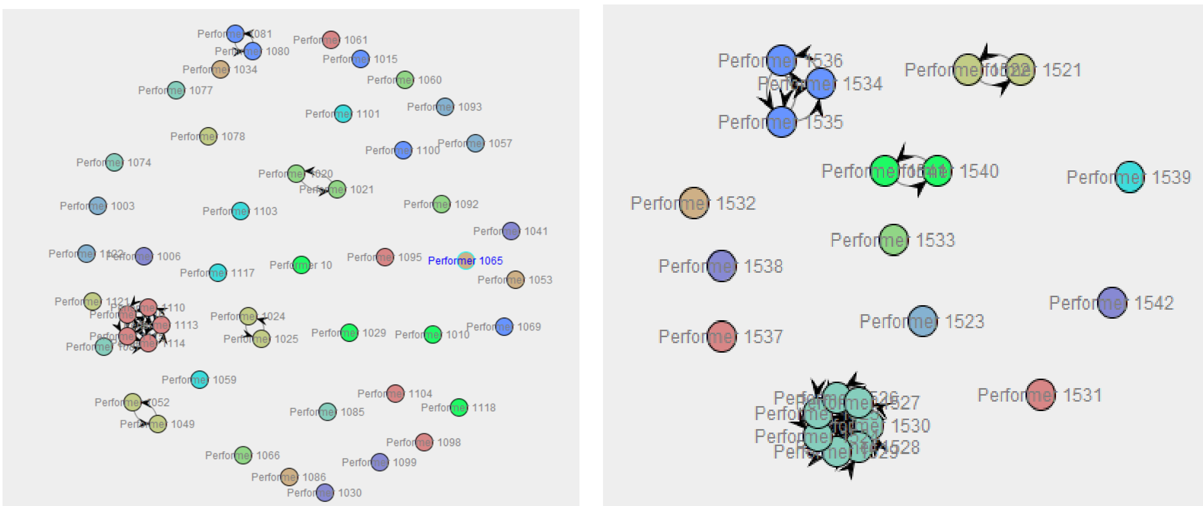


Figure 5.11: Social network of Similar-Task among resources in Radiology

By analyzing the event log, it was observed that performers in the red cluster in Figure 5.11 consistently engaged in the same activity, such as conducting CT scans. Similarly, the green cluster in Figure 5.11 comprised performers who performed the same activity, such as computed radiography (CR). Three activities are exclusively executed by specific individual resources: "MR" is performed solely by performer 1523, "DX" by resource 1521, and "MG" by resource 1522.

Finally, we look up the working together of social network miner which shows resources often work together in the same case. As shown in the Figure 5.12, all resources connected indicating sometimes working together on a case. In this study, we identified a social network reflecting collaborative work among resources in both diagnosis and radiology departments as shown in Figure 5.12. While most resources handled tasks for individual patients independently, some collaborated. This suggests that while individual performance is essential, there are critical instances of teamwork that contribute to more comprehensive patient management. The visual representation in Figure 5.12 reinforces the idea that understanding these collaborative patterns can inform strategies for improving workflow efficiency and enhancing outcomes in healthcare settings.

For example, in the diagnosis log, performers 1601, 1074, and 1078 worked individually on particular patients as shown in Figure 5.12(a). Conversely, performers 1110, 1113, 1114, and 1115 collaborated on the same patient. In the radiology department, nine resources worked independently as shown in Figure 5.12(b). However, other resources, such as performers 1524, 1525, 1526, 1527, 1528, 1529, and 1530, collaborated on the same patient.



(a) (b)

Figure 5.12: Social network of working together among resources in (a) diagnosis and (b) radiology

5.6. Design Requirement of the Framework

The experimentation analysis analyses conducted at Tikur Anbessa Specialized Hospital informed the design of the healthcare improvement framework by revealing critical insights into control flow, performance, and organizational dynamics. Control flow analysis identified bottlenecks that necessitated holistic integration of healthcare services, while performance analysis highlighted gaps in key metrics, emphasizing the need for continuous improvement through ongoing evaluation with process mining techniques. Additionally, organizational analysis tailored the framework to TASH's unique operational goals and patient demographics, ensuring relevance and effectiveness. Finally, a literature review uncovered gaps in existing research, underscoring the necessity for a comprehensive approach to healthcare integration and process optimization. Thus, this study investigates and establishes essential design requirements for constructing such a framework. These requirements ensure the framework is robust, scalable, and capable of addressing the complexities of healthcare processes. Key design requirements include:

- **Holistic Integration:** Combining healthcare service domains including healthcare systems and processes, which integrates input for delivering healthcare service such as guidelines and standards, infrastructure, socio-cultural factors, financial systems, and technology with process mining techniques like process discovery, conformance checking, and enhancement.
- **Continuous Improvement:** Utilizing process mining techniques to facilitate ongoing assessment and enhancement of healthcare processes. This approach involves systematically analyzing event data to identify inefficiencies, monitor performance metrics, and uncover areas for improvement. By continuously evaluating healthcare process, TASH can implement targeted strategies that enhance service delivery, optimize resource allocation, and ultimately improve patient care. This iterative process ensures that healthcare systems remain responsive and adaptive to changing needs and standards.
- **Customization for TASH:** Adapting the framework to meet the specific needs of TASH. This involves a thorough analysis of TASH's operational goals, patient demographics, and existing workflows to ensure that the framework aligns with its strategic vision. Key considerations include modifying processes, integrating relevant technologies, and incorporating specific guidelines that reflect the hospital's standards and practices.
- **Addressing Literature Gaps:** Developing a comprehensive framework not previously detailed in existing research. This involves identifying specific areas where previous studies have fallen short or where critical insights are lacking, particularly in the context of healthcare integration and process optimization. By conducting a thorough literature review, we have identified these gaps and established a framework that offers new perspectives and solutions.

5.7. Proposed Framework for Improving Healthcare Service

The proposed framework, illustrated in Figure 5.13, is designed to improve healthcare services using process mining techniques. It addresses patient expectations, organizational needs, and the dynamic healthcare environment. The components of the framework are grounded in empirical findings derived from our study and delineates key components essential for enhancing healthcare delivery. The framework is structured around several interconnected systems and processes, each addressing specific aspects of healthcare improvement. The proposed framework integrates the healthcare service domain with the process mining domain to enhance overall healthcare services.

- **Healthcare Service Domain**: Includes healthcare service inputs (such as guidelines and standards, infrastructure, socio-cultural and human systems, financial systems, and technology), healthcare systems and processes.
- **Process Mining Domain**: Comprises process discovery, conformance checking, and enhancement, analyzing models through control flow analysis, performance analysis, and social network analysis

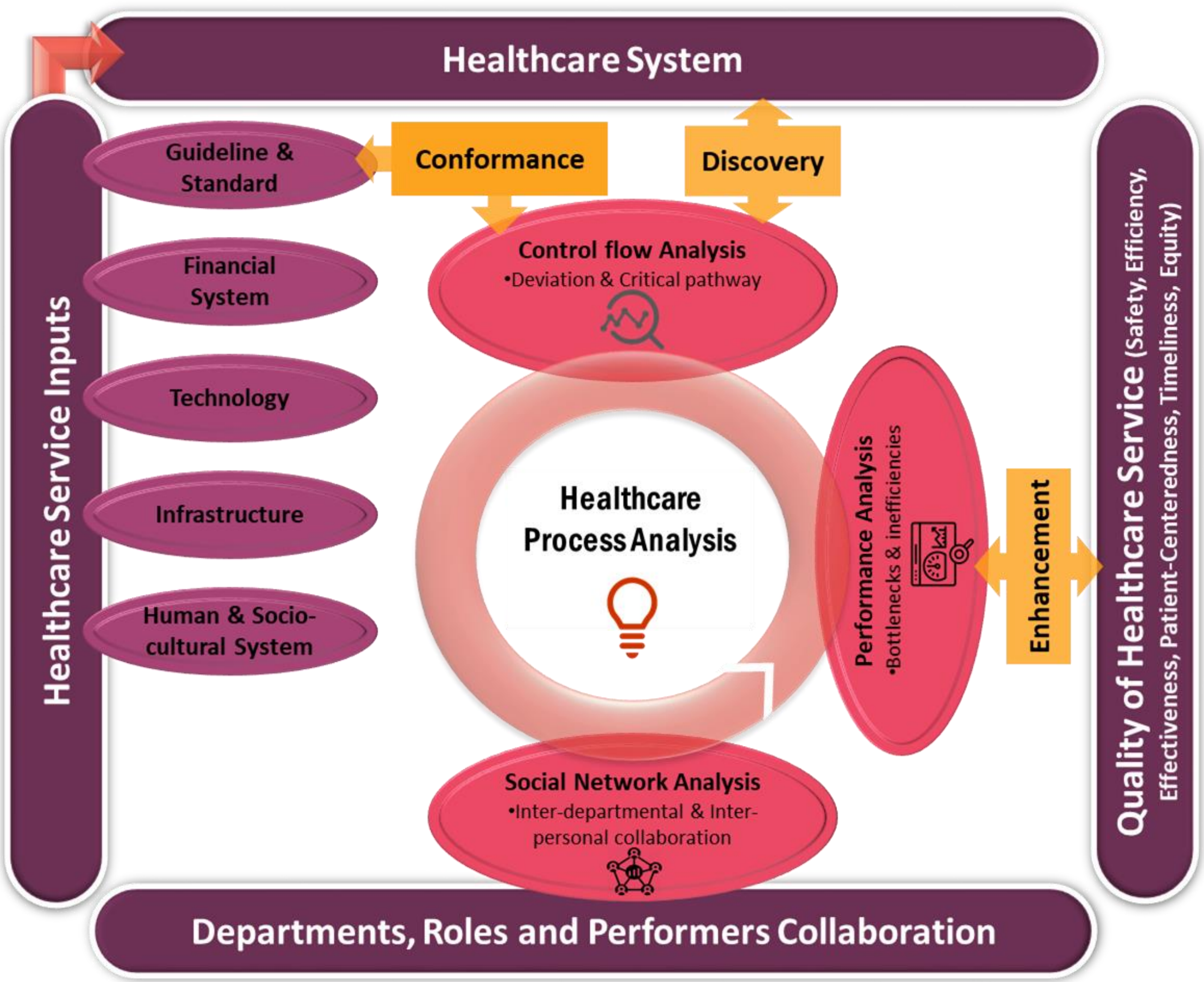


Figure 5.13: The proposed healthcare service improvement framework

Healthcare System: it represents the integration of various **inputs** essential for delivering healthcare services. It includes:

- **Guidelines & Standards**: Our study revealed that the laboratory test turnaround times exceeded WHO benchmarks, highlighting the need for rigorous comparison of existing practices against established standards. This emphasizes the importance of adhering to clinical protocols and regulatory compliance to ensure consistent, evidence-based care.
- **Infrastructure:** Findings indicated significant inadequacies in the physical and organizational infrastructure, which hindered the delivery of quality services. Addressing these gaps is critical for fostering an environment conducive to effective healthcare delivery.
- **Human & Socio-Cultural System:** The results underscored the impact of collaboration among healthcare performers on service quality. A robust socio-cultural framework that promotes teamwork and cultural competency is essential for improving patient engagement and care outcomes.
- **Financial System:** Our analysis identified imbalances in resource allocation as a significant inefficiency. By strategically directing funding to under-resourced areas, we can optimize the financial structure and improve overall service delivery.
- **Technology:** The study highlighted the role of advanced Health Information Systems (HIS) in enhancing service efficiency. However, challenges related to data quality and the incomplete implementation of these systems must be addressed to fully leverage technology for quality improvement.

Healthcare Process:

- **Clinical Processes:** The analysis of hematology treatment pathways revealed critical paths and alternative options followed by patients. This insight is crucial for refining clinical workflows to enhance patient care.

- **Operational Processes:** Day-to-day activities, such as scheduling and resource allocation, were examined to ensure seamless healthcare facility operations. Streamlining these processes is vital for maintaining efficiency and quality.

- **Support Processes:** Auxiliary functions, including human resources and supply chain management, support primary healthcare activities. Strengthening these support systems will enhance overall operational effectiveness.

Process mining

- **Discovery:** We employed process mining techniques to map the actual hematology treatment processes, identifying bottlenecks and inefficiencies. This discovery phase is fundamental for gaining insights into real-world practices.
- **Conformance:** We employed conformance techniques to evaluate the alignment between the discovered process models and the event log, measuring their fitness. Additionally, this approach allows us to assess compliance between the reference process model and the actual processes in practice.
- **Enhancement:** Based on our findings, we provided actionable recommendations to the hospital aimed at improving service quality through targeted process enhancements.

Process Analysis: Integrating process mining techniques into the healthcare service improvement framework enables a data-driven approach to optimizing processes.

- **Control Flow Analysis:** We conducted a thorough analysis of the hematology treatment process, identifying deviations and bottlenecks that impede efficiency. This analysis facilitates a more streamlined approach to patient care.
- **Performance Analysis:** Our evaluation of waiting times and resource utilization offered critical insights into operational efficiency, informing strategies for improvement.
- **Social Network Analysis:** By exploring stakeholder interactions, we identified key influencers and collaboration patterns among performers, which can enhance communication and optimize team dynamics.

The Departments, Roles, and Performers Collaboration: It focuses on enhancing communication and cooperation among various departments, roles, and individual performers within a healthcare organization. This collaboration is essential for improving patient outcomes, streamlining processes, and fostering a culture of continuous improvement.

The quality of healthcare service: The framework aims to enhance healthcare service quality through process mining by focusing on six key areas:

- **Safety**: Minimizing risks and harm to patients.
- **Efficiency**: Optimizing resource use for cost-effective care.
- **Effectiveness**: Delivering evidence-based care for the best outcomes.
- **Patient-Centeredness**: Respecting and addressing individual patient preferences and needs.
- **Timeliness**: Reducing wait times and delays in care.
- **Equity**: Ensuring consistent care quality regardless of personal characteristics such as gender, ethnicity, location, and socio-economic status.

5.8. Evaluation of the Proposed Framework

Evaluation is a vital part of the research process. According to [76], evaluating an artifact is necessary to demonstrate its effectiveness, utility, and efficiency. This evaluation process can be iterative, aiming to refine the framework and ensure its quality, thereby addressing real-world business problems. IT artifacts can be assessed based on several quality attributes such as functionality, comprehensiveness, consistency, accuracy, performance, reliability, usability, and organizational [76]. Various methods can be employed for evaluation, including observational, analytical, experimental, testing, expert validation, and descriptive approaches [76].

Before evaluating the proposed framework, it was crucial to present the findings of the study to the purposely selected evaluation respondents. This presentation covered how applying process mining in healthcare effectively identified bottlenecks, inefficiencies, and deviations in existing processes. Building on these findings, we explained the development of the framework, detailing its key components and how each is designed to address specific challenges within healthcare services. We also discussed the potential improvements the framework offers, emphasizing its capacity to enhance service delivery, optimize resource allocation, and ultimately contribute to better patient outcomes. This comprehensive overview aimed to ensure that respondents of evaluation are well-informed and prepared to provide insightful feedback during the evaluation process.

To demonstrate the presentation and evaluation of the proposed framework, two department heads from the hematology and radiology departments, six senior healthcare professionals from these departments, and four IT professionals were carefully selected through purposive sampling. In total, twelve participants, with extensive professional experience in management, medicine, and information technology, were involved in the presentation and evaluation process. The proposed framework was evaluated using expert validation and descriptive methods. Questionnaires, adapted from the evaluation criteria recommended by [76], were distributed to evaluators. The criteria included attributes such as organizational fit, comprehensiveness, reliability, clarity, correctness, and usability. Additionally, the evaluation checklist was adapted from Tigist (2018), cited in [39], and deemed appropriate for this study. The evaluation consisted of 10 question items (see Annex C), assessed using a Likert scale, and grouped into five basic categories. The scales are strongly disagree, disagree, neutral, agree, and strongly agree and their assigned weights are 1, 2, 3, 4, 5 respectively. The mean value was used to determine the central tendency of the framework. Before conducting the actual evaluation, the survey underwent a Cronbach's alpha test to check the reliability or internal consistency of the instrument, achieving a value of approximately 0. 747 as shown in

Table 5.4. According to [86], the value of Cronbach's alpha (0. 747) indicates that the survey is acceptable since it is greater than 0.7.

Table 5.4: Reliability statistics for the evaluation survey

Reliability Statistics	
Cronbach's Alpha	N of Items
.747	10

Based on the response of the experts response (see annex D), descriptive statistics shows, the evaluation of the proposed healthcare improvement framework reveals a strong positive reception in several critical areas with the overall mean value 4.2, as shown in Table 5.5. it indicates a strong tendency towards agreement, suggesting that respondents generally feel positively about the statements being evaluated. It's close to the upper end of the scale, reflecting that most participants are more than just neutral, they are inclined to agree with the proposed framework.

Respondents rated the organization and presentation highly suitable for TASH, with a mean score of 4.5, indicating a well-structured and appropriately tailored framework. The objectives of the framework are also clear, and its content is deemed highly relevant and scalable, demonstrating strong alignment with organizational needs and the ability to adapt to various scales of implementation. Moreover, the framework is perceived as easy to apply, with respondents agreeing on its potential to improve efficiency and effectiveness within the organization. Each of these aspects also scored a mean of 4.5, reflecting consistent and strong agreement among respondents. The low standard deviations in these areas further underscore the consensus and positive outlook towards the framework's applicability and benefits.

Table 5.5: Mean and standard deviation of the framework evaluation survey

Descriptive Statistics					
Items	N	Minimum	Maximum	Mean	Std. Deviation
The framework is comprehensive in terms of coverage	12	3.00	4.00	3.4167	.51493
The organization and presentation of the framework is suitable for TASH	12	4.00	5.00	4.5000	.52223
The objective of the framework is clear	12	4.00	5.00	4.5000	.52223
The content of the proposed framework is complete	12	3.00	4.00	3.6667	.49237
The content of the proposed framework is relevant	12	4.00	5.00	4.5000	.52223
The content of the proposed framework is clear	12	3.00	4.00	3.4167	.51493
The content of the proposed framework is scalable	12	4.00	5.00	4.5000	.52223
The proposed framework is easy to be applicable	12	4.00	5.00	4.5000	.52223
The applicability of the proposed framework can improve efficiency and effectiveness	12	4.00	5.00	4.5000	.52223
The implementation of the proposed framework fits with the organization problem	12	4.00	5.00	4.5000	.52223
Average				4.2	0.517784

However, there are areas for improvement, particularly in the comprehensiveness and clarity of the content, which received slightly lower ratings with a mean of 3.4167. These ratings suggest that while the framework covers many necessary aspects, there might be gaps that need addressing, and the content could be simplified or clarified further. Enhancing these areas will likely increase overall satisfaction and effectiveness, ensuring the framework meets the comprehensive needs of the organization and its stakeholders.

5.9. Discussion

This research addresses the three research questions by thoroughly exploring suitable process mining techniques for healthcare, identifying critical pathways, deviations, and bottlenecks, and demonstrating how insights can improve healthcare services.

To determine the optimal process mining technique for healthcare, we conducted a comparative assessment of heuristic mining and inductive mining methods, evaluating their respective process models. The inductive miner demonstrated higher performance in most of the process model evaluation metrics. Based on these findings, the inductive miner emerges as a more suitable choice for dynamic environments like healthcare organizations, in comparison to the heuristic miner.

This study investigates critical pathways, deviations, and bottlenecks in healthcare business processes, focusing on hematology patients. The analysis provides a comprehensive view of care flows, highlighting activity frequencies and significant interconnections. It also facilitates visualizations of patient pathways and activity durations, revealing bottlenecks and inefficiencies. Social network analysis further elucidates interdepartmental collaboration and workflow dynamics by examining work transfers and patterns of resource specialization.

The expert evaluation result shows the proposed framework is the best to improve healthcare service by using extracted insights in healthcare process. It combines process mining techniques with a focus on patient needs and healthcare system components to identify and address inefficiencies. The positive framework evaluation results and identified areas for improvement highlight the framework's potential to improve service delivery and patient satisfaction, making it a valuable tool for optimizing healthcare processes.

CHAPTER SIX

CONCLUSION AND RECOMMENDATION

The final chapter of this study includes several essential components. It starts with a conclusion that assesses the strengths and weaknesses of the study's findings. Following this, it presents recommendations based on these findings, emphasizing their practical implications. The chapter further explores the research's contributions to the field, and proposes potential directions for future research.

6.1. Conclusion

This research is aimed at applying process mining techniques within the healthcare context of TASH, with the objectives of gaining insights, identifying bottlenecks, and optimizing resource allocation. The process model encompasses various stages, including data extraction, preprocessing, process discovery, model evaluation, analysis from multiple perspectives, and transfer the result. This methodological framework is well-aligned with the study's goals, facilitating a comprehensive exploration of the underlying processes.

A notable strength of the study lies in the careful selection of research methods and tools, by a thorough review of existing literature. The utilization of the ProM process mining framework offered versatility and comprehensiveness, enabling the application of diverse techniques such as heuristic mining, inductive mining, and social network mining. These techniques facilitated a nuanced analysis, contributing to the depth and richness of findings.

The study provides significant insights into the healthcare processes at TASH. The analysis revealed that out of 3,804 patients, the majority (1,863) began their treatment with a laboratory request, underscoring the critical role of laboratory tests in guiding subsequent treatments. The most common pathway for hematology patients begins with a laboratory request, followed by a laboratory test, then a hematology diagnosis, and finally a prescription, highlighting this is the critical pathway of processes. However, the study identified notable discrepancies between the frequency of laboratory requests and the completion of laboratory tests, suggesting inefficiencies, including instances where test results were not properly recorded in the system. The prolonged average duration of 32 days for laboratory tests, significantly exceeding World Health Organization (WHO) benchmarks, was identified as a serious bottleneck, attributed to resource allocation imbalances. In our comparative analysis of the heuristic and inductive miner models, we found that while the heuristic miner achieved a marginally higher generalization score (0.99929 vs. 0.99561), the inductive miner outperformed in key metrics, including fitness (0.9759 vs. 0.9708), precision (0.47878 vs. 0.43736), and simplicity. These advantages suggest that the inductive miner is a more suitable model for our purposes. Furthermore, Rashed et al. [41] reported a fitness score of 0.9477 for the inductive miner, highlighting that our findings (0.9759) demonstrate a notable improvement in performance. Social Network Analysis (SNA) revealed strong interactions among various departments and roles within the hospital, with some tasks being performed independently and others collaboratively, particularly in the diagnosis and radiology departments. The study's proposed framework for improving healthcare services through process mining was positively received, with an overall mean evaluation score of 4.2, validated through expert feedback and achieving a Cronbach's alpha value of 0.747, indicating acceptable reliability. These findings underscore the complexity of healthcare processes and the need for continuous improvement through integrated systems.

The study encountered certain challenges, particularly in getting ready made data for process mining, data extraction and preprocessing. Extracting data from disparate sources across multiple departments posed complexities, necessitating integration into a unified database. Preprocessing was also challenging due to the lack of process-aware data, requiring the transformation into an event log format using various techniques such as aggregation, temporal approach and heuristic filtering.

6.2. Research Contribution

This research makes several significant contributions to the field of process mining in healthcare. Firstly, the research proposes tailored process mining methodologies that are suitable for both healthcare professionals with limited technical expertise and technical experts in process mining who may lack a deep understanding of clinical contexts. These methodologies are designed to be adaptable, allowing researchers to apply them across different industries with similar socio-economic and cultural contexts. This versatility enhances the potential impact of the methodologies beyond the healthcare sector.

Secondly, it involves the publication of a dataset derived from various database systems with different formats. The event log created from this dataset is tailored for process mining algorithms, making it accessible and useful for researchers focused on process mining techniques. This contribution provides a valuable resource for the academic community, enabling further studies and advancements in the field.

Finally, this research contributes significantly to the field of healthcare service improvement, by designing and developing a comprehensive framework that integrates healthcare service domains with process mining techniques. The proposed framework aims to enhance healthcare services by addressing patient expectations, organizational needs, and the dynamic nature of healthcare environments.

6.3. Recommendation for Practice

Based on the analysis, the following recommendations are proposed to enhance the efficiency, accuracy, and quality of healthcare delivery:

- Integrate the IWKET and MedWeb systems into a unified hospital information system (HIS) to eliminate manual data transfer, minimizing errors and saving time while ensuring seamless data flow across departments.
- Adopting Process-Aware Information Systems (PAIS) to streamline operations, reduce inefficiencies, and enhance patient care through real-time process monitoring and optimization.
- Extend the hospital's information system to cover inpatient services, as it currently the system only covers outpatients, to ensure the entire patient journey is captured and analyzed.
- Implement the proposed healthcare improvement framework to optimize healthcare processes, enhance patient satisfaction, and improve the overall quality and efficiency of services.
- Implement thorough training programs for healthcare staff on the new integrated HIS and improvement framework. Develop a strong change management strategy to address resistance and ensure a smooth transition to the new system.
- Create continuous feedback mechanisms from healthcare providers and patients to iteratively improve systems and processes. Define and regularly review performance metrics to assess the impact of the integrated HIS and process mining on healthcare quality, efficiency, and patient satisfaction.

6.4. Future Research Directions

Future research should focus on addressing the limitations of this study to improve the validity and comprehensiveness of the findings. Specifically, obtaining direct access to event logs from process-aware hospital information systems would be crucial. Such access would eliminate the need for extensive preprocessing, thereby reducing potential data quality issues and biases, and ensuring a more accurate and reliable dataset for process mining.

Expanding the implementation of healthcare information systems to include inpatient care should also be considered. This would allow for the collection of data necessary to apply process mining techniques to critical inpatient healthcare processes, providing deeper insights into areas needing improvement.

Conducting a simulation study of the proposed healthcare improvement framework is recommended to evaluate its potential impact on process efficiency, patient satisfaction, and overall service quality in a controlled environment before implementation.

REFERENCES

[1] E. Rojas, J. Munoz-Gama, M. Sepúlveda, and D. Capurro, "Process mining in healthcare: A literature review," *J. Biomed. Inform.*, vol. 61, pp. 224–236, 2016, doi: 10.1016/j.jbi.2016.04.007.

[2] R. S. Mans, W. M. P. Van der Aalst, and R. J. B. Vanwersch, *Process mining in healthcare: evaluating and exploiting operational healthcare processes*. SpringerInternational Publishing, pp. 1–26, 2015.

[3] Á. Rebuge and D. R. Ferreira, "Business process analysis in healthcare environments: A methodology based on process mining," *Inf. Syst.*, vol. 37, no. 2, pp. 99–116, 2012.

[4] P. PE, "The challenge of complexity in health care," *Br. Med. J.*, vol. 323, pp. 625–628, 2001.

[5] E. H. Bradley, S. Pallas, C. Bashyal, P. Berman, and L. Curry, "Developing strategies for improving health care delivery: Guide to Concepts, Determinants, Measurement, and Intervention Design," *World Bank, Washington, DC*, vol. 59885, 2010.

[6] T. Grisold, J. Mendling, M. Otto, and J. vom Brocke, "Adoption, use and management of process mining in practice," *Bus. Process Manag. J.*, vol. 27, no. 2, pp. 369–387, 2021, doi: 10.1108/BPMJ-03-2020-0112.

[7] W. M. P. Van Der Aalst, "Distributed process discovery and conformance checking," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7212 LNCS, pp. 1–25, 2012, doi: 10.1007/978-3-642-28872-2_1.

[8] W. Van Der Aalst, "Process mining: Overview and opportunities," *ACM Trans. Manag. Inf. Syst.*, vol. 3, no. 2, pp. 1–17, 2012, doi: 10.1145/2229156.2229157.

[9] P. A. Ganz, M. J. Hassett, and D. C. Miller, "Challenges and opportunities in delivering high-quality cancer care: a 2016 update," *Am. Soc. Clin. Oncol. Educ. B.*, vol. 36, pp. e294–e300, 2016.

[10] P. Homayounfar, "Process mining challenges in hospital information systems," in *2012 federated conference on computer science and information systems (FedCSIS)*, IEEE, 2012, pp. 1135–1140.

[11] J. Munoz-Gama *et al.*, "Process mining for healthcare: Characteristics and challenges," *J. Biomed. Inform.*, vol. 127, p. 103994, 2022.

[12] W. M. P. Günther, C. W., & Aalst, van der, "Fuzzy mining - adaptive process simplification

based on multi- perspective metrics Fuzzy Mining – Adaptive Process Simplification Based on Multi-perspective Metrics," in *5th International Conference on Business Process Management , Brisbane, Australia*, 2007, pp. 328–343. doi: 10.1007/978-3-540-75183-0.

[13] A. Partington, M. Wynn, S. Suriadi, C. Ouyang, and J. Karnon, "Process mining for clinical processes: a comparative analysis of four Australian hospitals," *ACM Trans. Manag. Inf. Syst.*, vol. 5, no. 4, pp. 1–18, 2015.

[14] T. Habte, Y. Tsige, and A. Cherie, "Survey on the quality of care standards in a nursing/midwifery training hospital at tikur anbessa specialized hospital," *Adv. Med. Educ. Pract.*, vol. 11, pp. 763–774, 2020, doi: 10.2147/AMEP.S260339.

[15] B. Adigo, D. Gulilat, and S. Kassa, "Delay in First Treatment Initiation and Associated Factors among Colorectal Cancer Patients at TikurAnbessa Specialized Hospital, Addis Ababa, Ethiopia: A Retrospective Cohort Study," *International Journal of Cancer Research & Therapy*, vol. 8, no. 4, pp. 155–163, 2023.

[16] Teklie, H., Engida, H., Melaku, B. et al., "Factors contributing to delay intensive care unit admission of critically ill patients from the adult emergency Department in Tikur Anbessa Specialized Hospital," *BMC Emerg Med.*, vol. 21, no. 1, Oct. 2021, doi: 10.1186/s12873-021-00518-z.

[17] F. N. Workneh, Z. M. Assen, K. Bizuwork, and B. D. Tolera, "Oral Health Awareness and Self-Care Practice among Diabetic Patients Attending," *Int. J. Diabetes Metab. Disord.*, vol. 8, no. 3, pp. 416–422, 2023.

[18] S. Zeleke, M. Anley, D. Kefale, and B. Wassihun, "Factors Associated with Delayed Diagnosis of Cervical Cancer in Tikur Anbesa Specialized Hospital, Ethiopia, 2019: Cross-Sectional Study," *Cancer Manag. Res.*, vol. 13, no. null, pp. 579–585, Jan. 2021, doi: 10.2147/CMAR.S285621.

[19] 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 *Biomedical Engineering Systems and Technologies: International Joint Conference, BIOSTEC 2008 Funchal, Madeira, Portugal, January 28-31, 2008 Revised Selected Papers 1*, Springer, 2009, pp. 425–438.

[20] W. Abo-Hamad, "Patient pathways discovery and analysis using process mining techniques: An emergency department case study," *Springer Proc. Math. Stat.*, vol. 210,

pp. 209–219, 2017, doi: 10.1007/978-3-319-66146-9_19.

[21] M. A. Grando, M. H. Schonenberg, and W. Van Der Aalst, "Semantic-based conformance checking of computer interpretable medical guidelines," in *Biomedical Engineering Systems and Technologies: 4th International Joint Conference, BIOSTEC 2011, Rome, Italy, January 26-29, 2011, Revised Selected Papers 4*, Springer, 2013, pp. 285–300.

[22] J. Poelmans, G. Dedene, G. Verheyden, H. Van der Mussele, S. Viaene, and E. Peters, "Combining business process and data discovery techniques for analyzing and improving integrated care pathways," in *Advances in Data Mining. Applications and Theoretical Aspects: 10th Industrial Conference, ICDM 2010, Berlin, Germany, July 12-14, 2010. Proceedings 10,* Springer, 2010, pp. 505–517.

[23] T. Sztyler, J. Völker, J. Carmona Vargas, O. Meier, and H. Stuckenschmidt, "Discovery of personal processes from labeled sensor data: An application of process mining to personalized health care," in *Proceedings of the International Workshop on Algorithms & Theories for the Analysis of Event Data: Brussels, Belgium, June 22-23, 2015*, CEUR-ws. org, 2015, pp. 31–46.

[24] G. Erdogan and A. Tarhan, "A goal-driven evaluation method based on process mining for healthcare processes," *Applied Sciences,* vol. 8, no. 6, p. 894, 2018.

[25] F. Caron, J. Vanthienen, K. Vanhaecht, E. Van Limbergen, J. De Weerdt, and B. Baesens, "Monitoring care processes in the gynecologic oncology department," *Comput. Biol. Med.*, vol. 44, pp. 88–96, 2014.

[26] L. Nyblade, M. Stockton, S. Travasso, and S. Krishnan, "A qualitative exploration of cervical and breast cancer stigma in Karnataka, India," *BMC Womens. Health*, vol. 17, no. 1, pp. 1–15, 2017, doi: 10.1186/s12905-017-0407-x.

[27] B. N. L. Calys-Tagoe, J. M. K. Aheto, G. Mensah, R. B. Biritwum, and A. E. Yawson, "Cervical cancer screening practices among women in Ghana: Evidence from wave 2 of the WHO study on global AGEing and adult health," *BMC Womens. Health*, vol. 20, no. 1, pp. 1–9, 2020, doi: 10.1186/s12905-020-00915-9.

[28] L. Alkire, S. Myrden, S. Köcher, and G. E. O'Connor, "Cultural Drivers of Health Engagement," *J. Int. Mark.*, vol. 31, no. 1, pp. 90–105, 2022, doi: 10.1177/1069031X221127938.

[29] L. Rogers, A. De Brún, and E. McAuliffe, "Defining and assessing context in healthcare

implementation studies: a systematic review," *BMC Health Serv. Res.*, vol. 20, no. 1, p. 591, 2020, doi: 10.1186/s12913-020-05212-7.

[30] G. T. A.-T. T.- Keppel, "Design and analysis : a researcher's handbook." Prentice Hall Englewood Cliffs, N.J., Englewood Cliffs, N.J., 1991. doi: LK - https://worldcat.org/title/646880748.

[31] M. Bozkaya, J. Gabriels, and J. M. Van Der Werf, "Process diagnostics: A method based on process mining," *Proc. - Int. Conf. Information, Process. Knowl. Manag. eKNOW 2009*, no. 1, pp. 22–27, 2009, doi: 10.1109/eKNOW.2009.29.

[32] M. Cho, M. Song, and S. Yoo, "A systematic methodology for outpatient process analysis based on process mining," in *Asia Pacific Business Process Management: Second Asia Pacific Conference, AP-BPM 2014, Brisbane, QLD, Australia, July 3-4, 2014. Proceedings 2*, Springer, 2014, pp. 31–42.

[33] S. Suriadi, R. S. Mans, M. T. Wynn, A. Partington, and J. Karnon, "Measuring patient flow variations: A cross-organisational process mining approach," in *Asia Pacific Business Process Management: Second Asia Pacific Conference, AP-BPM 2014, Brisbane, QLD, Australia, July 3-4, 2014. Proceedings 2*, Springer, 2014, pp. 43–58.

[34] S. Yoo *et al.*, "Assessment of hospital processes using a process mining technique: Outpatient process analysis at a tertiary hospital," *Int. J. Med. Inform.*, vol. 88, pp. 34–43, 2016, doi: 10.1016/j.ijmedinf.2015.12.018.

[35] C. Alvarez *et al.*, "Discovering role interaction models in the Emergency Room using Process Mining," *J. Biomed. Inform.*, vol. 78, pp. 60–77, 2018.

[36] F. Mannhardt and D. Blinde, "Analyzing the trajectories of patients with sepsis using process mining," in *RADAR+ EMISA 2017*, CEUR-ws. org, 2017, pp. 72–80.

[37] P. Rattanavayakorn and W. Premchaiswadi, "Analysis of the social network miner (working together) of physicians," in *2015 13th International Conference on ICT and Knowledge Engineering (ICT & Knowledge Engineering 2015)*, IEEE, 2015, pp. 121–124.

[38] A. J. M. M. Weijters, W. M. P. van der Aalst, and A. K. A. de Medeiros;, "Process Mining with the HeuristicsMiner Algorithm," *Beta Work. Pap.*, vol. 166, 2006.

[39] S. Bereket, "Big Data Governance Framework for Ethio Telecom: MSc Thesis," *Addis Ababa University*, Jan, 2022.

[40] J. C. A. M. Buijs, B. F. Van Dongen, and W. M. P. Van Der Aalst, "Quality dimensions in

process discovery: The importance of fitness, precision, generalization and simplicity," *Int. J. Coop. Inf. Syst.*, vol. 23, no. 1, pp. 1–39, 2014, doi: 10.1142/S0218843014400012.

[41] A. H. M. Rashed, N. E. El-Attar, D. S. Abdelminaam, and M. Abdelfatah, *Analysis the patients' careflows using process mining*, vol. 18, no. 2 February. 2023. doi: 10.1371/journal.pone.0281836.

[42] B. W. Van Der Aalst, "Process Mining," 2012, doi: 10.1145/2240236.2240257.

[43] A. Hakim, "Improving Healthcare Systems through Process Mining," *Proc. - 2020 23rd IEEE Int. Multi-Topic Conf. INMIC 2020*, pp. 4–7, 2020, doi: 10.1109/INMIC50486.2020.9318159.

[44] P. M. Ginter, W. J. Duncan, and L. E. Swayne, *The strategic management of health care organizations*. John Wiley & Sons, 2018.

[45] A. Stefanini, D. Aloini, R. Dulmin, and V. Mininno, "Linking Diagnostic-Related Groups (DRGs) to their Processes by Process Mining.," *HEALTHINF*, vol. 5, pp. 438–443, 2016.

[46] "Background of Tikur Anbessa Hospital | College of Health Sciences." Accessed: Aug. 13, 2024. [Online]. Available: https://www.aau.edu.et/chs/tikur-anbessa-specialized-hospital/background-of-tikur-anbessa-hospital/

[47] F. A. Yansen Mandacan, "Process mining in multidisciplinary healthcare," *Masters Thesis, Utr. Univ. MSc*, vol. 5, no. 1, 2021.

[48] S. C. Mathews, M. J. McShea, C. L. Hanley, A. Ravitz, A. B. Labrique, and A. B. Cohen, "Digital health: a path to validation," *NPJ Digit. Med.*, vol. 2, no. 1, p. 38, 2019.

[49] M. Pinho and A. Pinto Borges, "The views of health care professionals and laypersons concerning the relevance of health-related behaviors in prioritizing patients," *Heal. Educ. Behav.*, vol. 46, no. 5, pp. 728–736, 2019.

[50] B. Tibeme, H. Shahriar, and C. Zhang, "Process mining algorithms for clinical workflow analysis," in *SoutheastCon 2018*, IEEE, 2018, pp. 1–6.

[51] U. Ronellenfitsch, E. Rössner, J. Jakob, S. Post, P. Hohenberger, and M. Schwarzbach, "Clinical pathways in surgery—should we introduce them into clinical routine? A review article," *Langenbeck's Arch. Surg.*, vol. 393, pp. 449–457, 2008.

[52] K. Kirchner, N. Herzberg, A. Rogge-Solti, and M. Weske, "Embedding conformance checking in a process intelligence system in hospital environments," 2012, pp. 126–139.

[53] L. M. Napolitano, "Standardization of perioperative management: Clinical pathways,"

Surg. Clin. North Am., vol. 85, no. 6, pp. 1321–1327, 2005, doi: 10.1016/j.suc.2005.10.011.

[54] P. Lillrank and M. Liukko, "Standard, routine and non-routine processes in health care.," *Int. J. Heal. care Qual. Assur. Inc. Leadersh. Heal. Serv.*, vol. 17, no. 1, pp. 39–46, 2004, doi: 10.1108/09526860410515927.

[55] H. Maulidina, "Unsupervised Abstraction for Reducing the Complexity of Healthcare Process Models," *PhD Desertation, Univ. Leeds, Sch. Comput.*, p. 240, 2019.

[56] G. Riz, E. A. P. Santos, and E. D. F. R. Loures, "Process mining to knowledge discovery in healthcare processes," *Adv. Transdiscipl. Eng.*, vol. 4, pp. 1019–1028, 2016, doi: 10.3233/978-1-61499-703-0-1019.

[57] G. Kukreja and S. Batra, "Analogize process mining techniques in healthcare: Sepsis case study," *4th IEEE Int. Conf. Signal Process. Comput. Control. ISPCC 2017*, vol. 2017-Janua, pp. 482–487, 2017, doi: 10.1109/ISPCC.2017.8269727.

[58] H. B. Klasky and O. Ozmen," *Process mining in healthcare-a case study for the corporate data warehouse of the veterans affairs office*," no. September. 2019. [Online]. Available: https://www.osti.gov/biblio/1649520%0Ahttps://www.osti.gov/servlets/purl/1649520

[59] K. Toth, K. Machalik, G. Fogarassy, and A. Vathy-Fogarassy, "Applicability of process mining in the exploration of healthcare sequences," *IEEE 30th Jubil. Neumann Colloquium, NC 2017*, vol. 2018-Janua, pp. 151–155, 2018, doi: 10.1109/NC.2017.8263273.

[60] W. M. P. Van der Aalst, *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. 2011. doi: 10.1007/978-3-030-78829-2_15.

[61] W. Van Der Aalst *et al.*, "Process Mining Manifesto," In: Daniel, F., Barkaoui, K., Dustdar, S. (eds) Business Process Management Workshops. BPM 2011. Lecture Notes in Business Information Processing, vol 99, pp. 169–194, 2012, Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-28108-2_19.

[62] S. Van De, "Process Mining in Healthcare Mining for Cost and (Near) Incidents," *(Ph.D. thesis), Eindhoven Univ. Technol. Eindhoven*, 2015.

[63] W. M. P. Van Der Aalst, H. A. Reijers, and M. Song, "Discovering social networks from event logs," Comput. Support. Coop. Work, vol. 14, no. 6, pp. 549–593, 2005, doi: 10.1007/s10606-005-9005-9.

[64] W. Van der Aalst, T. Weijters, and L. Maruster, "Workflow mining: Discovering process models from event logs," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 9, pp. 1128–1142,

2004.

[65] A. V. Vidyapeetham, M. Campus, A. V. Vidyapeetham, M. Campus, A. V. Vidyapeetham, and M. Campus, “An Approach to Fuzzy Process Mining to Reduce Patient Waiting Time in a Hospital,” pp. 1–6, 2017.

[66] I. G. B.A and Univ. -Prof. Dr. Matthias Fink, “PROCESS MINING IN HEALTHCARE, Master’s Thesis,JOHANNES KEPLER UNIVERSITY LINZ,Austria,” 2022.

[67] A. Adtiansyah, “Aligning observed and modeled behavior. [Phd Thesis 1 (Research TU/e / Grad TU/e), Mathmatics and Computer Science]. Technische Universiteit Eindhoven.,” 2014.

[68] A. Adriansyah, J. Carmona, B. F. Van Dongen, and W. M. P. Van Der Aalst, Measuring precision of modeled behavior. 2015. doi: 10.1007/s10257-014-0234-7.

[69] World Health Organization, Organisation for Economic Co-operation and Development, and The World Bank, *Delivering quality health services: a global imperative for universal health coverage*. Geneva, Switzerland: World Health Organization, 2018. Available: https://www.who.int/publications/i/item/9789241513906.

[70] S. J. J. Leemans, D. Fahland, and W. M. P. Van Der Aalst, “Discovering block-structured process models from event logs - A constructive approach,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7927 LNCS, pp. 311–329, 2013, doi: 10.1007/978-3-642-38697-8_17.

[71] A. K. A. De Medeiros, A. J. M. M. Weijters, and W. M. P. Van Der Aalst, “Genetic process mining: A basic approach and its challenges,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 3812 LNCS, no. task C, pp. 203–215, 2005, doi: 10.1007/11678564_18.

[72] W. Zhao and X. Zhao, “Process Mining from the Organizational Perspective Á Social network analysis Á,” pp. 701–709, doi: 10.1007/978-3-642-54924-3.

[73] W. M. P. Van Der Aalst and M. Song, “Mining Social Networks : Uncovering Interaction Patterns in Business Processes,” pp. 244–245, 2004.

[74] D. R. Ferreira, “Performance Perspective BT - A Primer on Process Mining: Practical Skills with Python and Graphviz,” D. R. Ferreira, Ed., Cham: Springer International Publishing, 2020, pp. 47–64. doi: 10.1007/978-3-030-41819-9_4.

[75] M. Song and W. M. P. Van der Aalst, “Towards comprehensive support for organizational

mining,” *Decis. Support Syst.*, vol. 46, no. 1, pp. 300–317, 2008.

[76] Hevner AR, March ST, Park J, Ram S, “Design science in information systems research,” MIS quarterly, vol. 28, no. 1, pp. 75-105, 2004.

[77] W. M. P. Van Der Aalst, “Process Mining : Discovering and Improving Spaghetti and Lasagna Processes,” no. c, 2011.

[78] M. L. V. E. B, X. Lu, S. J. J. Leemans, and W. M. P. Van Der Aalst, “PM 2 : A Process Mining Project Methodology,” pp. 297–313, 2015, doi: 10.1007/978-3-319-19069-3.

[79] E. De Roock and N. Martin, “Process mining in healthcare – An updated perspective on the state of the art,” *J. Biomed. Inform.*, vol. 127, no. January, p. 103995, 2022, doi: 10.1016/j.jbi.2022.103995.

[80] E. Batista and A. Solanas, “Process mining in healthcare: A systematic review,” *2018 9th Int. Conf. Information, Intell. Syst. Appl. IISA 2018*, pp. 1–6, 2019, doi: 10.1109/IISA.2018.8633608.

[81] T. G. Erdogan and A. Tarhan, “Systematic Mapping of Process Mining Studies in Healthcare,” *IEEE Access*, vol. 6, pp. 24543–25567, 2018, doi: 10.1109/ACCESS.2018.2831244.

[82] S. Agostinelli, F. Covino, G. D’Agnese, C. De Crea, F. Leotta, and A. Marrella, “Supporting governance in healthcare through process mining: A case study,” *IEEE Access*, vol. 8, pp. 186012–186025, 2020, doi: 10.1109/ACCESS.2020.3030318.

[83] M. Rovani, F. M. Maggi, M. De Leoni, and W. M. P. Van Der Aalst, “Declarative process mining in healthcare,” *Expert Syst. Appl.*, vol. 42, no. 23, pp. 9236–9251, 2015, doi: 10.1016/j.eswa.2015.07.040.

[84] O. Dogan, “Process mining for check-up process analysis,” *Iioabj*, vol. 9, no. 6, pp. 56–61, 2018.

[85] B. F. Van Dongen, A. K. A. De Medeiros, H. M. W. Verbeek, and A. J. M. M. Weijters, “The ProM Framework :A New Era in Process Mining Tool Support,” pp. 444–454, 2005.

[86] Tavakol M, Dennick R. “Making sense of Cronbach's alpha,” *International journal of Medical Education*, vol. 2, pp. 53-55, 2011

ANNEXES

Annex A: AAU Support Letter

አዲስ አበባ ዩኒቨርሲቲ
የተፈጥሮና የኮምፒዩቴሽናል ሳይንስ ኮሌጅ
የኢንፎርሜሽን ሳይንስ ት/ቤት
አዲስ አበባ፣ ኢትዮጵያ

Addis Ababa University
College of Natural and Computational Sciences
School of Information Science
Addis Ababa, Ethiopia

Date: October 27,2023
Ref No.ST/SIS/014/2023/16

To: Tikur Anbesa Specialized Hospital Clinical Service Director
Addis Ababa

Subject:- Student Mesele Awulachew

Dear Sir /Madam,

Student Mesele Awulachew (ID.No GSR/0335/15) is a graduate student at the School of Information Science, Addis Ababa University. He is currently conducting M.Sc. Thesis research under the title "Application of Process Mining in Healthcare: Case Study in Black lion Hospital."

I would like to thank you in advance for all the assistance that you would provide to the student.

To Department of Internal Medicine
For Ethical & methodological review and giving necessary support mce approved.
09/11/2023

With Regards

Dereje Teferi (PhD)
Head, School of Information Science

አዲስ አበባ ዩኒቨርሲቲ የኢንፎርሜሽን ሳይንስ ትምህርት ቤት SCHOOL OF INFORMATION SCIENCE ADDIS ABABA UNIVERSITY

14/11/23
To Dr. Addisu
for your review

Tel. +251-1-122-91-91 P.O.Box. 1176 Fax. +251-1-1239729

Annex B: Ethical Clearance

ADDIS ABABA UNIVERSITY COLLEGE OF HEALTH SCIENCES
አዲስ አበባ ዩኒቨርስቲ ጤና ሳይንስ ኮሌጅ
INTERNAL MEDICINE DEPARTMENT (IRB)
የውስጥ ደዌ ህክምና ክፍል
Department Review Board

ANNEX3

IRB's Decision

Meeting No: 08/23
Protocol number: 104//23

Meeting Date: 31/08/2023

Protocol Title: "Applying Process Mining for Improving Healthcare Systems: Case study to Tkur Anbessa Hospital		
Principal Investigator:	**. Mesele Awulachew,**	
Institute:	College of Health Sciences, AAU Department of Internal Medicine	
Elements Reviewed	√ Attached ☐ Not attached	
Review of Revised Application ☐ Yes ☐ No	Date of Previous review:	
Decision of the meeting:	√ Approved ☐ Approved with Recommendation ☐ Resubmission ☐ Disapproved	

I. Elements approved
1. Protocol Version No: 2
2. Protocol Version Date:
3. Informed consent Version No. 2
4. Informed Consent Version Date:

II. Obligations of the PI
1. Should comply with the standard international & national scientific and ethical guidelines
2. All amendments and changes made in protocol and consent form needs IRB approval
3. The PI should report SAE within 10 days of the event
4. End of the study, including manuscripts and thesis works should be reported to the IRB
5. The PI should report non-compliance and unanticipated events

III. To NERC ☐

Institution Review Board (IRB) Approval: Period from **December 1, 2023 – January 30, 2023**

Follow up report expected in not applicable

2 Months √___ 6 Months ____ 9 Months _____ 10 years ___________

Chairperson, IRB
Dr. Addisu Melkie

Signature ___________
Date:22/11/2023

A. A. UNIVERSITY FACULTY OF MEDICINE
Department of Internal Medicine
ADDIS ABABA ETHIOPIA

Annex C: Proposed Framework Evaluation Survey

ADDIS ABABA UNIVERSITY

COLLEGE OF NATURAL AND COMPUTATIONAL SCIENCE

SCHOOL OF INFORMATION SCIENCE

Dear Sir/Madam,

I am a graduate student at Addis Ababa University, pursuing a Master of Science degree in Information System. I am conducting a research study titled "Applying Process Mining for Improving Healthcare System at Tikur Anbessa Specialized Hospital," as part of my final thesis project.

This questionnaire aims to evaluate a proposed healthcare improvement framework, specifically its comprehensiveness, clarity, completeness, correctness, and applicability. This framework seeks to enhance the healthcare system at Tikur Anbessa Specialized Hospital, and we believe the results could also benefit other hospitals with similar contexts and environments.

To achieve this, I am requesting your participation in completing a questionnaire designed to gather your expert insights and feedback on the proposed framework. Your responses will be invaluable in assessing the framework and ensuring its effectiveness and relevance in real-world settings. Participation is voluntary and will take approximately 20 minutes to complete. Rest assured that all information you provide will be kept confidential and used solely for academic purposes.

Thank you very much for your time and participation.

Sincerely,

Mesele Awulachew
mesele.awulachew@aau.edu.et

Please select your answers by putting (X) sign on the scale ranging strongly Agree through strongly Disagree in the appropriate space provided

	Strongly Disagree	**Disagree**	**Neutral**	**Agree**	**Strongly Agree**
General					
The proposed framework is comprehensive in terms of coverage					
The organization and presentation of the framework is suitable for TASH					
The objective of the framework is clear					
The content of the proposed framework is complete					
Regarding the content of the Framework					
The content of the proposed framework is relevant					
The content of the proposed framework is clear					
The content of the proposed framework is Scalable					
Regarding utility and applicability of the framework					
The proposed framework is easy to be applicable					
The applicability of the proposed framework can improve efficiency and effectiveness					
The implementation of the proposed framework fits with the organization problem					

Annex D: Summary of experts response

	Strongly Disagree (1)	Disagree (2)	Neutral (3)	Agree (4)	Strongly Agree (5)
General					
The proposed framework is comprehensive in terms of coverage	0	0	7	5	0
The organization and presentation of the framework is suitable for TASH	0	0	0	6	6
The objective of the framework is clear	0	0	0	6	6
The content of the proposed framework is complete	0	0	4	8	0
Regarding the content of the Framework					
The content of the proposed framework is relevant	0	0	0	6	6
The content of the proposed framework is clear	0	0	7	5	0
The content of the proposed framework is Scalable	0	0	0	6	6
Regarding utility and applicability of the framework					
The proposed framework is easy to be applicable	0	0	0	6	6
The applicability of the proposed framework can improve efficiency and effectiveness	0	0	0	6	6
The implementation of the proposed framework fits with the organization problem	0	0	0	6	6