Utilizing Process Mining with Large Synthetic Datasets: A Case Study on Chemotherapy Treatment for Breast Cancer

Fatin Ramli *

aHealth Economics and Decision Science, School of Medicine and Population Health, University of Sheffield, Sheffield, United Kingdom.

bFaculty of Industrial Management, Universiti Malaysia Pahang, Gambang, Malaysia.

Abstract

This paper presents a case study on integrating process mining using a large synthetic dataset of breast cancer cases named Simulacrum. The aim was to analyze the chemotherapy treatment process through process mining. However, challenges were encountered due to the absence of experts for guidance. This study discusses the challenges faced, potential opportunities, and the essential requirements for successfully integrating process mining with significant datasets.

1. INTRODUCTION

Healthcare data analytics is a promising field that investigates the intricacies of patient care, treatment efficacy, and the development of health systems. The other aspect of this analysis is to acknowledge that not all people are alike. Every patient has his or her own unique qualities such as medical history, genetic makeup, lifestyle factors, and personal preferences that determine how they respond to treatment significantly. In the course of examining healthcare data, one looks at large quantities of information from numerous sources including electronic health records (EHRs), medical imaging, genetic data and patient recommended outcomes. The understanding of patient heterogeneity is what in part gives rise to various treatments. Healthcare providers strive to personalize treatments for individual patients based on such considerations as disease severity, comorbidities, previous treatments and potential side effects. This approach not only boosts the effectiveness of treatments but also minimizes risks while improving patient care [1-3].

Process mining is an emerging field that includes extracting information from event logs for discovering, monitoring and improving real processes [4]. In this case study we aim to incorporate process mining using Simulacrum dataset which is a synthetic dataset mimicking breast cancer cases and chemotherapy treatments [5]. Lamine, Fontanili, Di Mascolo, and Pingaud [6] introduce an approach for process reengineering in Emergency Call Services management by combining process mining and discrete event simulation techniques. One of the key strengths of this approach is its efficiency in quickly discovering process models and obtaining useful figures and statistics for incoming call processing. In Zhou, Wang, and Li [7], an application of advanced methods was developed to improve healthcare processes. The utilization of event logs and process mining to uncover a process model for clinical care processes represents a reliable approach. This method provides a data-driven and systematic way to analyze and optimize healthcare workflows, moving beyond reliance on observational data.

Kovalchuk et al. [8] introduced an approach by combining process mining techniques with discrete-event simulation (DES) in healthcare. One of the significant strengths of this approach is its ability to detect and assess diversity in patient flow and classes of clinical pathways. This allows for a more realistic and detailed simulation of patient flow, contributing to a comprehensive understanding of healthcare processes.

2. METHODOLOGY