*Proc. of the European Safety and Reliability Conference (ESREL 2026)*
*Edited by* José C. Matos, Paulo B. Lourenço, Michael Beer, Daniel Oliveira,
Edoardo Patelli, Hélder S. Sousa, Rui A. Silva and Monica Santamaria
 *Published by* Research Publishing, Singapore.
doi: 10.3850/ESREL2026061419_esrel26-p27785-cd

# Privacy-Preserving Process Mining for Clinical Pathway Analysis and Classification

Francesco Vitale

*Department of Electrical Engineering and Information Technology, University of Naples Federico II, Italy.*
*E-mail: francesco.vitale@unina.it*

Nicola Mazzocca

*Department of Electrical Engineering and Information Technology, University of Naples Federico II, Italy.*
*E-mail: nicola.mazzocca@unina.it*

Simone Guarino

*Department of Engineering, University Campus Bio-Medico of Rome, Italy.*
*E-mail: s.guarino@unicampus.it*

Process mining in healthcare offers significant benefits, including modeling patients' clinical pathways, identifying the load on healthcare infrastructure, and extracting useful statistics for analytics. However, these innovations pose privacy risks, such as the leakage or misuse of patient data, which could potentially disclose medical conditions and cause psychological distress or economic harm. In this context, privacy-preserving process mining is essential in order to maintain these advantages while safeguarding individual privacy. This paper proposes a framework that enables the analysis, classification and statistical study of patients' clinical pathways, while protecting against privacy threats. The framework supports various privacy-preserving techniques, such as introducing controlled noise through differential privacy, while integrating well-known process discovery and conformance checking algorithms for clinical pathways analysis and classification. We evaluate the effectiveness of the framework on an open-access dataset covering various diseases. The results demonstrate that we can achieve high classification accuracies for different clinical pathways while preventing malicious attempts to infer sensitive patient information, resulting in an effective trade-off between data utility and privacy guarantees.

*Keywords*: Healthcare, electronic health records, process mining, privacy-preserving techniques, differential privacy.

## 1. Introduction

The increasing automation of healthcare processes has improved their analysis and facilitated decision-making (Penev et al., 2024). What is more, automation is exponentially growing, as new machine learning-based approaches are being developed (Li et al., 2025). In this paper, we focus on a class of approaches grounded in process mining, a family of techniques that bridges data science and process science (van der Aalst and Carmona, 2022). These techniques can build process-based characterizations of clinical pathways, which are healthcare procedures that track patients' clinical courses while being treated for specific diseases (Vitale and Mazzocca, 2025). By modeling the clinical pathways followed by the patients throughout their clinical course, clinicians can monitor their progress, assess the severity of the disease, and predict or even prevent future outcomes (Eili et al., 2025).

Despite the remarkable progress in digitalizing healthcare processes, careful attention must be paid to the handling and processing of patients' data. Healthcare applications must prioritize patient privacy and data protection, as the leakage or misuse of sensitive medical information can lead to a loss of patient trust, reputational damage, and significant economic consequences arising from legal actions against healthcare institutions (Pool et al., 2024). This concern has become even more critical with the introduction of stringent data protection frameworks such as the General Data Protection Regulation (GDPR), which imposes strict requirements on how personal and health-related data are collected, stored, processed, and shared (de Kok et al., 2023).

In this paper, we are investigating the ability of analyzing historical clinical pathways and classifying the clinical pathways of new patients while preserving their privacy. Existing privacy-preserving approaches aim at anonymizing event data by a number of techniques, including encryption (Kazemian and Helfert, 2023), generalization into equivalent classes (Rafiei and van der Aalst, 2021), or noise addition (Fahrenkrog-Petersen et al., 2020). While these procedures effectively preserve the privacy of clients' data, they impact control-flow analyses as they may invalidate correct activity sequencing (Pika et al., 2020). In turn, it becomes difficult to correctly classify new clinical pathways. Our proposal allows maintaining data utility and preserving privacy by

- accounting for event data anonymization in characterizing and classifying clinical pathways;
- enforcing privacy-preserving transformations to new data in untrustworthy processing zones.

We demonstrate our framework's capabilities in classifying clinical pathways using Synthea (Walonoski et al., 2020), a public synthetic dataset that includes the generation of COVID-19 patients' clinical courses affected by different health complications.

The rest of the paper is structured as follows. Section 2 reports related works on privacy-preserving process mining techniques and their use in healthcare; Section 3 details the proposed framework; Section 4 details the two datasets above and reports the results; and Section 5 concludes and presents future works.

## 2. Related works

### 2.1. *Clinical pathway analysis through process mining*

Process mining involves capturing process-based characterizations of event data (process discovery), verifying deviations from prescriptive behavior (conformance checking), and enhancing the characterization with other process perspectives, such as the elapsed time of the events and their cost (process enhancement) (van der Aalst and Carmona, 2022). The main advantage of process mining lies in its ability to build interpretable process models and provide explainable process-based diagnoses across different application domains (Vitale et al., 2025).

Notably, process mining has been extensively used to characterize healthcare processes, including treatment procedures and management of hospital resources (Munoz-Gama et al., 2022). In this paper, we are interested in using process mining for clinical pathway analysis, which is particularly suited due to its process-based nature. For example, process mining has been adopted for modeling the patients' hospitalization in the presence of an incisional hernia (Phan et al., 2019), COVID-19 disease management in the presence of comorbidities (Cuendet et al., 2022), and treatment paths for heart failure (Beyel et al., 2024).

### 2.2. *Privacy-preserving process mining*

While existing literature has examined both the opportunities and challenges of applying process mining in the healthcare domain, greater emphasis is required on preserving patients' privacy in compliance with GDPR guidelines (Rafiei and van der Aalst, 2020). Privacy concerns extend beyond explicit identifiers, such as personal information and case IDs, to include activity sequences, which may reveal rare behavioral patterns that can be traced back to individual patients (Mannhardt et al., 2019). To address these risks, privacy-preserving techniques have been developed to anonymize event data used in process mining. A central challenge in privacy-preserving process mining lies in balancing robust privacy guarantees with the preservation of data utility (Rafiei and van der Aalst, 2020).

A widely adopted class of privacy-preserving techniques is group-based anonymization based on, e.g., k-anonymity and l-diversity, aiming at generalizing event data into equivalence classes in order to mitigate the risk of disclosing information linked to individual cases (Rafiei and van der Aalst, 2021; Fahrenkrog-Petersen et al., 2023). Another prominent approach is based on $\epsilon$-differential privacy, a probabilistic framework that protects individual records by injecting carefully

calibrated noise into the data (Mannhardt et al., 2019; Fahrenkrog-Petersen et al., 2020).

Despite the increasing interest in the use of process mining in the healthcare sector and the wide range of existing techniques for anonymizing event data, privacy-preserving clinical pathway analysis and classification remain largely underexplored in the literature. We aim to preserve the utility of this task while ensuring patient privacy through a framework that performs clinical pathway analysis on historical data within a trustworthy processing zone. The resulting models are then transferred for reuse on anonymized data within an untrustworthy processing zone, enabling the accurate classification of new patient cases with minimal loss in performance.

## 3. Framework

Our framework aims at maintaining the advantages of clinical pathway analysis and classification through process mining while ensuring the patients' privacy in high-risk environments. To this aim, the framework, shown in Fig. 1, splits the tasks and data of the approach into two trustworthy and untrustworthy processing zones.

### 3.1. *Trustworthy Processing Zone*

The goal of the trustworthy processing zone is to characterize **clinical pathways** and build a **privacy-aware classifier** able to classify patients' clinical pathways while tolerating their anonymized data collected in the untrustworthy processing zone.

First, historical data are obtained from reference electronic health records. In this paper, we consider having data encoded as simple event logs, which are collections of traces (van der Aalst and Carmona, 2022). An event log is denoted as $L \in \mathcal{B}(\mathcal{A}^*)$, where $\mathcal{A}$ is the set of treatment procedures, $\mathcal{A}^*$ is the set of possible traces that can be built with the treatment procedures in $\mathcal{A}$, and $\mathcal{B}(\mathcal{A}^*)$ is the universe of possible multisets of traces. A trace $\sigma \in \mathcal{A}^*$ is an ordered sequence of $k$ treatment procedures, i.e., $\sigma = \langle a_1, a_2, \ldots, a_k \rangle, a_i \in \mathcal{A}$. In the following, we will focus on the processing of a historical event log $L_{CP}$ of a given clinical pathway $CP$.

$L_{CP}$ is split into training and validation sublogs $L_{CP,Tr}, L_{CP,Val} \subset L_{CP}$. On the one hand, a process discovery algorithm is applied to the traces of $L_{CP,Tr}$ to obtain the process-based characterization of a clinical pathway $N_{CP}$, e.g., a Petri net (Vitale and Mazzocca, 2025). On the other hand, $L_{CP,Val}$ is processed through an anonymization process aimed at generating an anonymized validation event log $\tilde{L}_{CP,Val}$.

In this paper, we will focus on enforcing $\epsilon$-differential privacy, as this type of anonymization significantly impacts process mining results more than other anonymization approaches (Pika et al., 2020). $\epsilon$-differential privacy is defined as follows. Let $d$ and $d'$ be two neighboring datasets, differing in at most one individual record, and let $\mathcal{M}$ be a randomized mechanism with output range $\mathcal{R}$. The mechanism $\mathcal{M}$ satisfies $\epsilon$-differential privacy if, for all measurable subsets $S \subseteq \mathcal{R}$,

$$\Pr[\mathcal{M}(d) \in S] \leq e^{\epsilon} \cdot \Pr[\mathcal{M}(d') \in S].$$

The parameter $\epsilon > 0$ controls the privacy–utility trade-off: smaller values of $\epsilon$ yield stronger privacy guarantees by limiting how much the output distribution can change when a single individual's data is modified. The $\mathcal{M}$ mechanism has been implemented through SaCoFa, an algorithm that inserts noise into a trace-variant count by tuning the $\epsilon$ parameter; the lower $\epsilon$ is, the higher the noise injected into the original trace variants, hence the higher the privacy guarantee (Kirchmann et al., 2022).

The traces of $\tilde{L}_{CP,Val}$ can now be checked against $N_{CP}$ and generate a set of conformance-checking diagnoses $\mathcal{D}_{CP} \in \mathbb{R}^{|\tilde{L}_{CP,Val}| \times |\mathcal{A}|+1}$ (Vitale et al., 2025). These allow outlining the activities in the anonymized traces that fail to conform to the behavior captured in $N_{CP}$ and the fitness values of the traces. By systematically recording and accounting for these non-conformities, a privacy-aware classifier can be constructed. Such a classifier is designed to recognize clinical pathways even when their observed traces are misaligned with the reference behavior due to data anonymization. This objective can be achieved using machine learning techniques, such as support vector machines or

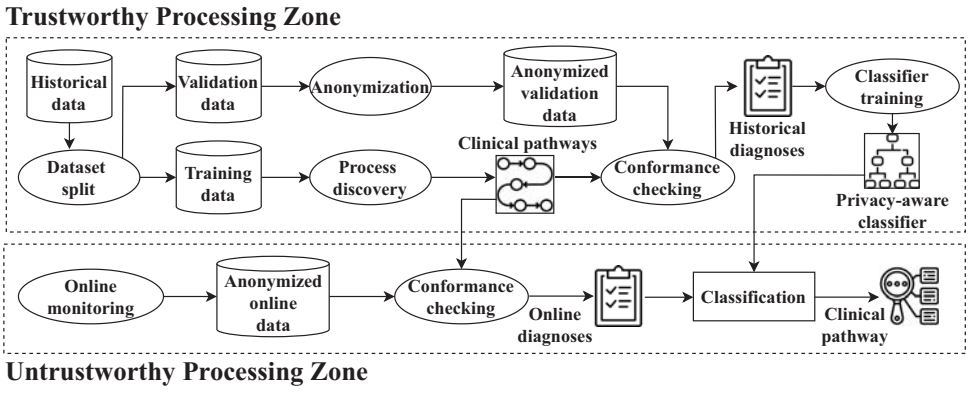


Fig. 1. The framework for privacy-aware analysis and classification of clinical pathways.

decision trees (Singh et al., 2022).

### 3.2. *Untrustworthy Processing Zone*

The goal of the untrustworthy processing zone is to perform online monitoring of patients' treatment traces and store their anonymized form in order to avoid disclosure of sensitive information. A similar approach has been proposed by Mannhardt et al. (2019), whose approach interfaces an untrusted environment through a privacy engine that anonymizes the event data through $\epsilon$-differential privacy. In contrast to their work, which focuses on a centralized privacy-preserving interface for event log analysis, our approach assumes that the untrusted processing zone may be realized by distributed nodes with high mobility and computational efficiency but limited security guarantees, a setting that is increasingly common in modern healthcare application scenarios (Rancea et al., 2024). The proposed just-in-time anonymization aims to reduce re-identification risk in line with GDPR principles and is informed by Recital 28, which states that "the application of pseudonymisation to personal data can reduce the risks to the data subjects concerned and help controllers and processors to meet their data-protection obligations."[a]

Given an anonymized online trace $\tilde{\sigma}$, this can be checked against $N_{CP}$ to obtain online diagnoses. These can be classified with the privacy-aware classifier, resulting in the identified clinical pathway.

Table 1. Number of traces for each sublog of the event logs extracted from the Synthea dataset.

| Event log | Training | Validation | Test |
|---|---|---|---|
| $L_{PE}$ | 52 | 26 | 26 |
| $L_{ARD}$ | 118 | 58 | 58 |
| $L_{SST}$ | 66 | 33 | 33 |
| $L_{AVP}$ | 258 | 128 | 128 |
| $L_{AB}$ | 233 | 115 | 115 |
| $L_{MNB}$ | 80 | 39 | 39 |

## 4. Experiments

### 4.1. *Dataset, techniques and metrics*

The Synthea dataset[b] records synthetic electronic health records collected from the realistic simulation of COVID-19 patient treatments (Walonoski et al., 2020). It includes several clinical pathways followed by patients with different COVID-19 complications. In our experiments, we extracted the COVID-19 patients' clinical pathways with Pulmonary Emphysema (PE), Acute Respiratory Distress syndrome (ARD), Streptococcal Sore Throat (SST), Acute Viral Pharyngitis (AVP),

[a] https://gdpr-info.eu/recitals/no-28/

[b] https://synthea.mitre.org/downloads

Acute Bronchitis (AB), and Malignant Neoplasm of Breast (MNB). We collected all the treatment procedures of each patients and built six event logs. Each event log was split into three sublogs, holding out 25% of testing traces to evaluate the classification ability of our framework and 25% of validation traces to extract the conformance-checking diagnoses and build the privacy-aware classifier. Table 1 reports the number of traces for each sublog.

In our experiments, we have used several well-known process discovery algorithms to evaluate their impact on the classification results: the Inductive Miner Infrequent (IMf), Integer Linear Programming Miner (ILP), and the Heuristics Miner (HM) (van der Aalst and Carmona, 2022). The algorithms have been configured with a high 90% noise tolerance to mitigate the "spaghetti" effect seen in healthcare processes (Bosson-Rieutort et al., 2025). As for classifier training, we have used eXtreme Gradient Boosting (XGBoost), a decision-tree-based machine learning approach for supervised learning (Chen and Guestrin, 2016). As mentioned in Section 3, we have used the SaCoFa algorithm for anonymizing validation and test traces. However, we have also implemented a non-privacy-aware variation of the classifier so that we could compare the benefit of anonymizing validation traces on the classifier's effectiveness.

To evaluate the performance of XGBoost, we employed the macro F1-Score (F1), which calculates the F1 independently for each class and then takes the unweighted average. Let $TP_i$, $FP_i$, and $FN_i$ denote the true positives, false positives, and false negatives for class $i$, respectively. The precision and recall for class $i$ are defined as:

$$\text{Precision}_i = \frac{\text{TP}_i}{\text{TP}_i + \text{FP}_i}$$

$$\text{Recall}_i = \frac{\text{TP}_i}{\text{TP}_i + \text{FN}_i}$$

$$\text{F1}_i = 2 \cdot \frac{\text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i}$$

The macro F1 $\frac{1}{C}\sum_{i=1}^{C} \text{F1}_i$ is the unweighted average of the F1 across all classes.

The experiments for each process discovery algorithm and $\epsilon$ value were repeated 5 times. The software is implemented using Python and uses several process mining and machine learning packages, including `pm4py` and `scikit-learn`. The code is available online on GitHub[c].

### 4.2. *Results and discussion*

Table 2 shows the F1 results achieved by the non-privacy-aware and privacy-aware XGBoost classifier for each process discovery algorithm and $\epsilon$-value combination. The table highlights that the privacy-aware classifier almost always achieves the best results compared to the non-privacy-aware one, demonstrating that using anonymized validation traces for classifier training improves the performance against the test set. This is further shown in Fig. 2, which highlights the side-by-side comparison of the trends followed by the privacy-aware and non-privacy-aware classifiers.

It is worth noting that the process discovery algorithm significantly impacts the classifier results regardless of its privacy awareness. This is to be expected, as different process discovery algorithms yield process models with different structures due to their representational bias, which influences the *assumptions* of the algorithms about the underlying process as control-flow relations are mined (van der Aalst, 2011). For example, the HM provides very low F1 figures for each $\epsilon$ value, whereas the IMf is able to achieve up to 0.908 F1 and 0.749 F1 for the privacy-aware and non-privacy-aware settings, respectively, at $\epsilon = 1.0$. Notably, the IMf is able to maintain 0.714 F1 for the strongest privacy setting, i.e., $\epsilon = 0.1$, using the privacy-aware classifier.

In conclusion, the results indicate that the proposed framework is able to maintain consistently high F1 scores ($\geq 0.714$) across all tested values of $\epsilon$ when using IMf, while achieving substantial improvements over the non–privacy-aware setting. Although these findings are promising and align with the underlying rationale of the framework, additional experimentation is required to further validate the results. In particular, fu-

[c] https://github.com/francescovitale/pppm_cp

Table 2. Mean macro F1-Scores achieved by the Non-Privacy-Aware (NPA) and Privacy-Aware (PA) XGBoost classifier for each process discovery algorithm and $\epsilon$-value combination. The bold figures indicate the best scores.

| $\epsilon$ | IMf | | ILP | | HM | |
|---|---|---|---|---|---|---|
| | PA | NPA | PA | NPA | PA | NPA |
| 1.0 | **0.908 ± 0.075** | 0.749 ± 0.098 | **0.697 ± 0.038** | 0.464 ± 0.047 | 0.564 ± 0.043 | **0.597 ± 0.073** |
| 0.75 | **0.892 ± 0.052** | 0.873 ± 0.069 | **0.696 ± 0.063** | 0.468 ± 0.041 | 0.561 ± 0.054 | **0.593 ± 0.072** |
| 0.5 | **0.848 ± 0.102** | 0.825 ± 0.145 | **0.647 ± 0.051** | 0.494 ± 0.039 | 0.537 ± 0.075 | **0.580 ± 0.063** |
| 0.25 | **0.825 ± 0.131** | 0.736 ± 0.100 | **0.681 ± 0.088** | 0.487 ± 0.088 | **0.539 ± 0.104** | 0.509 ± 0.094 |
| 0.1 | **0.714 ± 0.080** | 0.553 ± 0.125 | **0.671 ± 0.128** | 0.390 ± 0.070 | **0.490 ± 0.079** | 0.445 ± 0.067 |

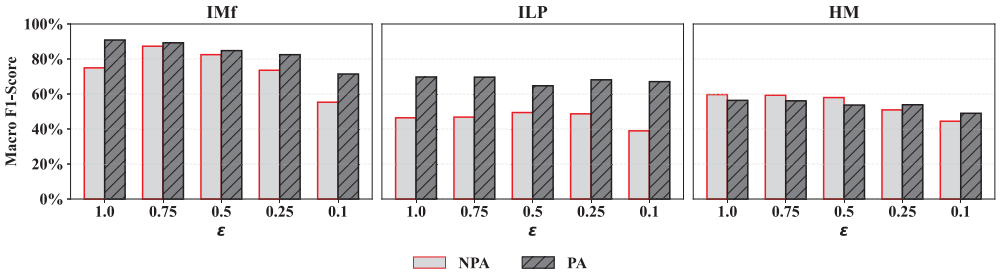


Fig. 2. Side-by-side comparison of the macro F1-Scores achieved by the Non-Privacy-Aware (NPA) and Privacy-Aware (PA) XGBoost classifier for each process discovery algorithm and $\epsilon$-value combination.

ture work should consider alternative privacy-preserving techniques (e.g., k-anonymity), conduct more extensive sensitivity analyses of the process discovery algorithms and classifiers, and evaluate the framework on real-world datasets to better address both internal and external validity threats.

## 5. Conclusion

Automatic analysis and modeling of clinical pathways is a crucial step in modern healthcare applications. In this context, process mining is particularly valuable due to its ability to draw process-based insights directly from data. However, it is mandatory to employ privacy-preserving frameworks to protect the clients' identities and prevent economic losses, in accordance with the guidelines of the GDPR. To this aim, we developed a privacy-preserving process mining framework for clinical pathway analysis and classification, which maintains the ability to interpret different clinical courses while properly classifying anonymized clients' data. The results showed that our framework is able to achieve satisfactory classification scores across different experimental settings, maintaining processing utility while guaranteeing $\epsilon$-differential privacy.

Although the method application highlighted promising results, it has several limitations, including: 1) using only $\epsilon$-differential privacy without extending to other anonymization frameworks; 2) not outlining a pre-processing pipeline to handle real-life electronic health records; 3) not providing partial trace support for incomplete and/or interrupted treatments; and 4) limited sensitivity analysis of the involved process mining and machine learning algorithms. Future works will aim to address these limitations and further extend the method to distributed environments, where ensuring data anonymization is more challenging due to data dispersion across different nodes. Finally, it is worth noting that integrating an explicit, guided pre-processing step into the method enables its application to a wider range of business processes where data anonymization is

a central requirement, including finance, telecommunications, and supply chain management. In these highly regulated domains, data anonymization is not merely a best practice but a strict necessity that ensures organizations can optimize operations without violating legal mandates or compromising user trust.

**Acknowledgement**

This work has been partially supported by the Spoke 9 "Digital Society & Smart Cities" of ICSC - Centro Nazionale di Ricerca in High Performance-Computing, Big Data and Quantum Computing, funded by the European Union - NextGenerationEU (PNRR-HPC, CUP: E63C22000980007).

## References

Beyel, H., M. Verket, V. Peeva, C. Rennert, M. Pegoraro, K. Schütt, W. van der Aalst, and N. Marx (2024). Process-Aware Analysis of Treatment Paths in Heart Failure Patients: A Case Study. In *Proceedings of the 17th International Joint Conference on Biomedical Engineering Systems and Technologies - HEALTHINF*, pp. 506–515. INSTICC.

Bosson-Rieutort, D., A. Langford-Avelar, J. Duc, and B. Dalmas (2025). Healthcare trajectories of aging individuals during their last year of life: application of process mining methods to administrative health databases. *BMC Medical Informatics and Decision Making 25*(1), 58.

Chen, T. and C. Guestrin (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794.

Cuendet, M. A., R. Gatta, A. Wicky, C. L. Gerard, M. Dalla-Vale, E. Tavazzi, G. Michielin, J. Delyon, N. Ferahta, J. Cesbron, et al. (2022). A differential process mining analysis of COVID-19 management for cancer patients. *Frontiers in oncology 12*, 1043675.

de Kok, J. W., M. Á. A. de la Hoz, Y. de Jong, V. Brokke, P. W. Elbers, P. Thoral, A. Castillejo, T. Trenor, J. M. Castellano, A. E. Bronchalo, et al. (2023). A guide to sharing open healthcare data under the general data protection regulation. *Scientific data 10*(1), 404.

Eili, M. Y., J. Rezaeenour, and M. H. Roozbahani (2025). Predicting clinical pathways of traumatic brain injuries (tbis) through process mining. *npj Digital Medicine 8*(1), 112.

Fahrenkrog-Petersen, S. A., H. van der Aa, and M. Weidlich (2020). Pripel: privacy-preserving event log publishing including contextual information. In *International Conference on Business Process Management*, pp. 111–128. Springer.

Fahrenkrog-Petersen, S. A., H. van der Aa, and M. Weidlich (2023). Optimal event log sanitization for privacy-preserving process mining. *Data & Knowledge Engineering 145*, 102175.

Kazemian, M. and M. Helfert (2023). A lightweight encryption method for privacy-preserving in process mining. In *2023 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech)*, pp. 0228–0233.

Kirchmann, H., S. A. Fahrenkrog-Petersen, M. Kabierski, H. van der Aa, and M. Weidlich (2022). Privacy-preserving process mining with pm4py. In *ICPM Doctoral Consortium/Demo*, pp. 85–89.

Li, Y., J. Tian, A. Xu, R. Greiner, J. Hayward, A. J. Greenshaw, and B. Cao (2025). Maturity framework for operationalizing machine learning applications in health care: Scoping review. *Journal of Medical Internet Research 27*, e66559.

Mannhardt, F., A. Koschmider, N. Baracaldo, M. Weidlich, and J. Michael (2019). Privacy-preserving process mining: Differential privacy for event logs. *Business & Information Systems Engineering 61*(5), 595–614.

Munoz-Gama, J., N. Martin, C. Fernandez-Llatas, O. A. Johnson, M. Sepúlveda, E. Helm, V. Galvez-Yanjari, E. Rojas, A. Martinez-Millana, D. Aloini, et al. (2022). Process mining for healthcare: Characteristics and challenges. *Journal of Biomedical Informatics 127*, 103994.

Penev, Y. P., T. R. Buchanan, M. M. Ruppert, et al. (2024). The Evolution of Health Information Technology for Enhanced Patient-Centric Care in the United States: Data-Driven Descriptive Study. *Journal of Medical Internet Research 26*, e59791.

Phan, R., V. Augusto, D. Martin, and M. Sarazin (2019). Clinical pathway analysis using process mining and discrete-event simulation: an application to incisional hernia. In *2019 winter simulation conference (WSC)*, pp. 1172–1183. IEEE.

Pika, A., M. T. Wynn, S. Budiono, A. H. Ter Hofstede, W. M. van der Aalst, and H. A. Reijers (2020). Privacy-preserving process mining in healthcare. *International journal of environmental research and public health 17*(5), 1612.

Pool, J., S. Akhlaghpour, F. Fatehi, and A. Burton-Jones (2024). A systematic analysis of failures in protecting personal health data: A scoping review. *International Journal of Information Management 74*, 102719.

Rafiei, M. and W. M. van der Aalst (2021). Group-based privacy preservation techniques for process mining. *Data & Knowledge Engineering 134*, 101908.

Rafiei, M. and W. M. P. van der Aalst (2020). Privacy-preserving data publishing in process mining. In *Business Process Management Forum*, Cham, pp. 122–138. Springer International Publishing.

Rancea, A., I. Anghel, and T. Cioara (2024). Edge computing in healthcare: Innovations, opportunities, and challenges. *Future Internet 16*(9).

Singh, P., M. Saman Azari, F. Vitale, F. Flammini, N. Mazzocca, M. Caporuscio, and J. Thornadtsson (2022). Using log analytics and process mining to enable self-healing in the internet of things. *Environment Systems and Decisions 42*(2), 234–250.

van der Aalst, W. (2011). On the representational bias in process mining. In *2011 IEEE 20th International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises*, pp. 2–7.

van der Aalst, W. M. P. and J. Carmona (2022). *Process Mining Handbook*. Cham, Switzerland: Springer.

Vitale, F. and N. Mazzocca (2025). Adaptive identification and modeling of clinical pathways with process mining.

Vitale, F., M. Pegoraro, W. M. van der Aalst, and N. Mazzocca (2025). Control-flow anomaly detection by process mining-based feature extraction and dimensionality reduction. *Knowledge-Based Systems 310*, 112970.

Walonoski, J., S. Klaus, E. Granger, D. Hall, A. Gregorowicz, G. Neyarapally, A. Watson, and J. Eastman (2020). Synthea™ Novel coronavirus (COVID-19) model and synthetic data set. *Intelligence-based medicine 1*, 100007.