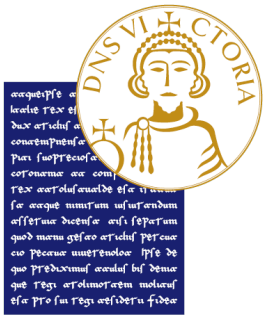

University of Sannio

Department of Engineering

Ph.D. in Information Technologies for Engineering

Process Mining in Healthcare: Maintenance of Data and Models

Ph.D. Coordinator:

Professor Massimiliano Di Penta

Ph.D. Student:

Antonella Madau

Supervisor:

Professor Lerina Aversano

Benevento, date

A te, Mamma,

che avresti voluto essere qui,

continuerò a portarti con me , sempre.

Abstract

This thesis investigates the complexities associated with the application of Process Mining (PM) and Machine Learning (ML) in the healthcare domain, with a specific focus on the long-term maintenance of data and models. Despite the transformative potential of these technologies in optimizing clinical pathways and operational efficiency, their adoption is frequently hindered by the high heterogeneity of medical data, the non-linear nature of clinical processes, and the performance degradation of predictive models over time, commonly known as concept drift.

To address these challenges, this research proposes an integrated framework that encompasses the entire data lifecycle, ranging from automated acquisition and semantic standardization to robust process discovery and adaptive predictive modelling. The primary contribution of this work lies in the development of a pipeline for the semantic reconciliation of heterogeneous and evolving healthcare data, complemented by a process-aware methodology designed to extract complex temporal and behavioral features from care pathways.

Furthermore, the proposal introduces a comprehensive governance and maintenance layer dedicated to continuous model monitoring, drift detection, and human-in-the-loop validation. Experimental results, validated on real-world clinical datasets, demonstrate that integrating a process-oriented perspective significantly enhances the accuracy, interpretability, and resilience of predictive analytics within healthcare environments.

Keywords: Process Mining, Healthcare Analytics, Machine Learning Maintenance, Concept Drift Detection, Clinical Pathways, Data Governance, Predictive Modelling, Evidence-based Medicine

Contents

Introduction

The digital transformation of modern healthcare systems has led to the widespread adoption of Electronic Health Records (EHRs), resulting in the accumulation of large volumes of heterogeneous clinical data. This growing availability of data creates unprecedented opportunities to apply data-driven approaches, such as Process Mining and Machine Learning, to support clinical decision-making, improve patient outcomes, and optimize hospital workflows.

Despite this potential, the practical application of these techniques in healthcare remains challenging. Clinical processes are characterized by high variability in patient pathways, complex organizational structures, and the continuous evolution of medical guidelines, all of which limit the effectiveness of traditional analytical methods when applied in isolation.

Within this context, this thesis identifies three critical research gaps. First, clinical data are often fragmented across heterogeneous information systems and require substantial manual effort to be transformed into reliable and semantically consistent event logs. Second, the intrinsic flexibility of clinical care frequently leads process mining techniques to produce overly complex and scarcely interpretable models. Third, predictive models deployed in real-world settings are subject to performance degradation over time due to evolving data distributions, highlighting the lack of formal governance and systematic maintenance strategies.

The objective of this research is to define a process-aware and adaptive methodological framework that ensures healthcare analytics are not only accurate at the time of deployment, but also reliable, transparent, and maintainable throughout their operational lifecycle.

CHAPTER

1 Background

1.1 Process Mining and Machine Learning in Healthcare

The combined application of process mining (PM) and machine learning (ML) is increasingly crucial for improving clinical pathways and managing healthcare information. The advancement of hospital digitalization, combined with the growing availability of data from heterogeneous systems such as electronic medical records, triage systems, laboratories, and monitoring devices, now enables process analysis with an unprecedented level of detail. All of these elements make an integrated approach essential [29, 32].

In this context, PM reconstructs and analyzes actual execution flows, highlighting variability, protocol deviations, operational inefficiencies, and complex behaviour patterns[5]. ML, in parallel, introduces predictive capabilities to anticipate clinical events, improve decision support, and identify anomalies in patient pathways early [24]. The combination of the two disciplines creates a powerful and complementary analytical ecosystem: process mining provides the structure of the process, its temporal dynamics, and its actual behaviour, while machine learning leverages this information to derive accurate and adaptive predictive models[38].

This synergy is particularly relevant in healthcare, where process variability is structural and depends on differences between patients, the continuous evolution of protocols, and the natural overlap of activities[31]. Clinical processes are complex, non-linear, and subject to constant change, making traditional static modelling approaches insufficient [34].

Studies show how PM and ML cooperate naturally: PM highlights gaps or discontinuities in processes, and ML can reconstruct missing activities or infer hidden states [37]. The structural complexity of logs is highlighted by PM, which becomes informative material for ML, exploiting it as a predictive basis. The time sequences produced by PM provide ML with key chronological features on which to build more accurate models[25]. Finally, process models derived from real-world execution are used by machine learning to predict clinical outcomes more accurately[38]. The combination of machine learning and machine learning not only allows us to describe what happens in healthcare processes but also to predict their effects, contributing to the development of robust analytics systems capable of supporting operational and clinical decisions in highly dynamic contexts[3].

1.1.1 Process Mining Fundamentals

Process Mining emerged as a fundamental discipline, acting as a strategic bridge between Business Process Management (BPM) and Data Science, with the primary goal of extracting objective knowledge and process insights directly from the analysis of execution data recorded in corporate information systems[3]. This approach bridges the gap between how processes are theoretically designed and how they are actually executed.

The foundation of every Process Mining technique is the Event Log, which describes the real, objective behaviour of a process and serves as a database for extracting new models, comparing actual execution with expected patterns, and identifying bottlenecks and deviations, the canonical and essential structure is the essential prerequisite for launching any type of analysis[5]. This structure is characterized by three key and interdependent components: the case identifier (Case ID), which uniquely groups all activities related to a single process instance (e.g., a single patient journey); the Activity label, which specifies the action or step performed (such as ”Initial Diagnosis” or ”Test Execution”); and the Timestamp, which provides a precise timestamp of the start or end of the activity, essential for defining the sequential order and calculating the durations of each process [3].

An event and its attributes are formally defined as follows:

1. **Definition 1 (Event)** *Let* $X_{Event} = \{x_1, \ldots, x_p\}$ *be a finite set of attributes (such as*

timestamp, activity type, case ID, duration, etc.), where $p \in \mathbb{N}$. *An event defined on* X_{Event} *is a set of p values, one for each attribute. Each event is uniquely determined by the combination of all its attribute values.*

2. **Definition 2 (Trace)** *Let T be a set of events. A trace* σ *is an ordered sequence of events from T:*

$$\sigma = \langle c_1, \ldots, c_n \rangle$$

where for each $i \in [1,n]$, $c_i \in T$. *Here, n is the length of the trace. The set of all traces over T is denoted as* T^*.

3. **Definition 3 (Event Log)** *Let T be a set of events. A log L over T is a non-empty set of traces over T:*

$$L = \{\sigma_1, \ldots, \sigma_m\}$$

where $m \in \mathbb{N}$, *and for each* $i \in [1,m]$, $\sigma_i \in T^*$. *The events within a given log are defined on the same set of attributes, although they may have different values.*

Starting from this structured data, Process Mining expands into three interconnected macro-areas (Figure 1.1): Process Discovery, Conformance Checking, and Process Enhancement.

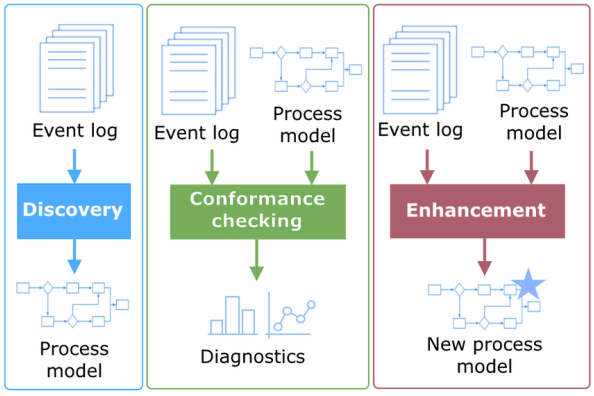

Figure 1.1: Macro-Areas of Process Mining

Process Discovery aims to automatically construct a representative process model (often a workflow or Petri net) directly from the logs[1], answering the question of how the process is actually executed. Conformance Checking verifies the degree of adherence between the

recorded execution and a normative or expected model [14] and is essential for identifying deviations, fraud, or rule violations [4]. Despite its importance, literature reviews show that it is currently less represented in healthcare research compared to process discovery [22]. Finally, Process Enhancement uses insights from the first two areas to improve the process, focusing on identifying and addressing operational inefficiencies, bottlenecks, or suboptimal path variants [31].

In the specific context of healthcare services, Process Discovery is the area that faces the most daunting challenges due to the intrinsic nature of clinical data[31]. Indeed, healthcare processes are clearly distinguished from those in other sectors by their intrinsic complexity: clinical event logs exhibit high dimensionality, heterogeneous information, and, above all, high behavioural variability due to the personalization of treatment pathways[35]. This variability exposes Process Discovery algorithms to various issues, such as strong sensitivity to noise, duplication of activities, and temporal variability. In-depth studies, such as those related to Time-Window Activity Aggregation, have shown that the presence of redundantly recorded activities can be misinterpreted as a cyclical and frequent loop in the model, distorting its true representation[11]. Furthermore, the multiplicity of legitimate clinical pathways and the potential temporal overlap of activities lead to the generation of a very large number of process variants, resulting in final models known as ”spaghetti models”: complex structures that are difficult to visualize, non-scalable, and nearly impossible for clinical analysts to interpret[31].

Due to these challenges, Process Mining has emerged as a successful tool in the clinical field, and the need to reduce the dimensionality and complexity of input event logs has become imperative. As described in complex methodologies [11], techniques such as stratified sampling allow the selection of log subsets that maintain a statistically faithful representation of the most significant process variants, while abstraction based on time or patterns aims to consolidate micro-sequential events into single, more significant macro-activities[37]. This need for rigorous preparation is widely documented in the literature as a primary requirement for successful process mining in the clinical setting [35].These preprocessing strategies are not optional, but necessary and essential: they highlight that the effective application of Process Mining to healthcare data requires specific and rigorous attention to data properties and the

adoption of advanced data preparation and maintenance mechanisms.

1.1.2 Machine Learning for Healthcare

In addition to flow analysis, Machine Learning (ML) has become a critical tool in healthcare, supporting areas such as diagnosis, prognosis, and patient monitoring by identifying complex patterns within high-dimensional clinical data. Within healthcare Process Mining (PM), Machine Learning (ML) serves a complementary function: PM structures and defines process behaviours while ML interprets and predicts these behaviours. This synergy forms the basis of hybrid PM+ML approaches,these are becoming the standard for advanced healthcare process analysis by enabling real-time operational decision-making and enhancing clinical flow management [25]. On the one hand, Machine Learning (ML) contributes to data quality improvement, which is essential for reliable process analyses. This is demonstrated by Suriadi et al. [36], which focuses on repairing missing activity labels in clinical event logs. In this context, predictive models and structural constraints derived from the process are combined to restore the original semantics of the traces, demonstrating how ML can be used not only to predict outcomes but also to remediate data imperfections that underpin process analyses.

On the other hand, Machine Learning extends the capabilities of PM by enabling predictive monitoring. Studies conducted on complex clinical datasets, such as those related to emergency medicine (e.g. MIMIC-IV), demonstrate how features extracted from patient pathways and temporal dynamics can feed models capable of anticipating clinical evolution [18]. In particular, deep learning architectures such as LSTM networks [13, 37] and multi-view approaches [33] have shown superior performance in predictive business process monitoring by capturing long-term dependencies in event sequences. The effectiveness of such models is strictly linked to the integration of temporal features: capturing the sequence and timing of events is in fact crucial to improve the ability to anticipate outcomes in critical contexts [38].

In conclusion, the integration between Process Mining and Machine Learning is a strategic element for supporting operational decisions and optimizing the management of clinical flows.

1.1.3 Process Mining in Clinical Contexts

The application of Process Mining (PM) in healthcare is crucial because it transforms data generated by information systems (such as Electronic Medical Records) into visual and quantifiable models of real-world workflows[5]. However, this context faces intrinsic challenges related to complex, highly dynamic, and rarely linear processes. Clinical pathways involve multiple actors, different skills, and activities that can be repeated, overlapping, or influenced by external events, generating significant inter-patient variability that significantly impacts the complexity of the extracted models.

Healthcare processes present specific characteristics compared to traditional processes,the accuracy of results depends heavily on the quality of the input datasets, and clinical data, generated by heterogeneous systems, frequently present semantic gaps or structural incompleteness, as demonstrated by work on log repair[35]. This missing data in event logs represents the most significant quality issue, potentially leading algorithms to generate inaccurate models or identify non-existent causal connections.

Clinical event logs are also distinguished by extreme characteristics that complicate their management and analysis: they are characterized by very high dimensionality and strong heterogeneity, due to significant inter-patient variability. This variability leads to the creation of an extremely high number of unique trace variants; consequently, managing very long traces and numerous variants, as highlighted by studies on sampling and aggregation, makes the direct application of standard PM techniques difficult and requires significant computing power. Added to this is the problem of noise, often caused by duplicate activities found in event sequences: this noise distorts the discovered pattern, increasing unnecessary complexity and reducing the reliability of compliance metrics.

Furthermore, temporal analysis introduces a further level of complexity, showing that clinical pathway and temporal context are closely intertwined. Execution patterns, in fact, can depend on external factors such as the time of day or the operational load of the department, a vital detail for predictive models. This complexity highlights the dual need to develop models that are both interpretable and scalable for real-world clinical use. An overly complex model, even if statistically accurate, would be ineffective if it were incomprehensible to healthcare

professionals and inapplicable to large volumes of real-time data.

Finally, evidence from studies conducted on real emergency room data demonstrates how the integration between process mining and machine learning can provide significant insights into patient journey management [18]. This hybrid approach highlights the ability of process mining to faithfully represent complex operational dynamics, concretely supporting clinical decisions. This synergy offers a methodological solution to address the intrinsic variability of healthcare systems [35]., ensuring the practical adoption of advanced analytics tools in hospital workflows.

In summary, although the literature highlights the synergistic value of Process Mining and Machine Learning, a methodological fragmentation emerges: existing studies tend to isolate data quality improvement (Log Repair) from long-term model maintenance (Model Maintenance). The challenge is not only to integrate these two disciplines, but to create an analytical ecosystem capable of self-correcting in the face of structural variability in clinical processes.

1.2 Healthcare Systems and Data Characteristics

Modern healthcare systems are extremely complex, characterized by a dense network of interactions between medical and nursing staff, patients, equipment, and administrative infrastructure. These systems operate in a highly critical environment, where real-time decisions and quality of care are top priorities. The digitalization of clinical processes, through hospital information systems and Electronic Health Records (EHR), has led to the generation of a massive amount of data. However, the very nature of the clinical workflow and the technologies employed give this data characteristics that make it particularly challenging for advanced analysis, challenging the direct application of traditional data analysis or process mining (PM) techniques[5].

The main critical issues can be traced back to three interconnected factors, clearly evident in the analysis of complex datasets such as MIMIC-IV which integrates data from multiple sources [30]:

- Structural and Semantic Heterogeneity: Each clinical or operational system was designed independently, with its own logic, purposes, and standards. For example, a laboratory produces highly structured data with specific encodings, while triage records events in text form. These differences concern not only the data format, but also the meaning attributed to the same concepts (incompatible nomenclatures, different scales). Consequently, the integration of these heterogeneous sources—such as EHRs, triage systems, laboratories, administrative systems, and monitoring devices—results in inconsistencies, ambiguities, and misalignments. These inconsistencies require significant standardization, semantic mapping, and harmonization of information flows to achieve a consistent representation. To address these challenges, it is essential to also consider the attributes present in clinical logs and the crucial role of metadata, which is essential for disambiguating labels and providing the necessary context for analysis.

- The Dynamism and Non-Linearity of Patient Pathways: Clinical evolution does not follow an orderly and repeatable progression, but depends on a succession of events that can abruptly alter the direction of the pathway. Clinical decisions emerge in relation to test results and vital signs monitored in real time, making the healthcare workflow a set of states that activate flexibly and reactively, forming a branched, sometimes recursive, pathway governed more by contingency than planning. The extracted data reflects this complexity, with long traces and a significant number of variants, resulting in extremely diverse behaviour across patients that requires specific approaches, such as those explored in stratified sampling and aggregation studies, to manage the high behavioural complexity.

- Inconsistent Healthcare Data Quality: This is a direct consequence of operational priorities and technical limitations of the systems. Recording is not the operator's primary objective, which inevitably leads to omissions, delays, and inaccurate entries. Added to this are technical issues that generate duplications, unsynchronized timestamps, and missing values (such as missing labels, the prevalence of which is demonstrated by work on task repair). The end result is a dataset that reflects the imperfection of the real context. Addressing these imperfections requires a systematic approach to identify log

patterns [36], followed by reconstruction techniques such as decision tree learning to mitigate the tendency of missing data [28]. Time also represents a critical and complex dimension: minimal differences in timestamps can significantly influence the order of events. Furthermore, time-window aggregation reveals that many clinical activities are repeated every few seconds or minutes, generating temporal noise that makes it difficult to discover an interpretable pattern. The importance of the temporal dimension is not only operational, but also diagnostic, as the timing of events is crucial for prognosis[38].

In conclusion, the heterogeneous, incomplete, and noisy nature of clinical data requires that the application of Process Mining and Machine Learning be preceded by accurate and robust data standardization and pre-processed strategies. Only by properly managing these issues can the full analytical potential of advanced techniques be exploited.

1.3 Challenges

The application of process mining in healthcare presents a series of structural challenges that arise from the very nature of clinical processes and the characteristics of the data produced by hospital information systems. Unlike other sectors, healthcare is characterized by highly heterogeneous operational flows, high inter-patient variability, and the presence of dynamic environments in which protocols, technologies, and clinical conditions change rapidly. These elements directly influence the quality of available data and the ability of models to accurately represent real-world processes [35].

A first critical issue concerns the heterogeneity of data sources. Clinical systems integrate information from different departments, each with its own recording methods, specific granularity, and semantic structures that are not always compatible. The consequence is the difficulty of building a coherent event log, in which activities are correctly aligned and integrated. This heterogeneity not only represents a technical obstacle but also profoundly impacts the discovery phase and the quality of the generated models, which risk reflecting data inconsistencies rather than actual process flows.

A second crucial challenge is the incompleteness of clinical logs. Healthcare data often

contains missing activities, incorrect labels, broken sequences, or inconsistent timestamps. Work on reconstructing missing activities has highlighted how these gaps compromise the semantics of the process and make it difficult to reconstruct reliable clinical pathways. Without repair strategies, process discovery is distorted, compliance assessment is unreliable, and predictive analytics are significantly degraded[36]. Adding to these challenges is the constant evolution of clinical protocols and organizational practices. Diagnostic and therapeutic pathways change over time in response to new clinical evidence, technological updates, or changes in workloads. Studies on temporal aggregation and temporal features clearly demonstrate how even minimal changes in activity execution methods produce observable variations in temporal patterns and process sequences. This phenomenon generates a natural "process drift"[38] that, if not identified and managed, compromises the ability of models to represent clinical reality.

A further element of complexity arises from the continuous transformation of the patient population. In hospital departments, particularly in emergency areas, seasonal, epidemic, or demographic variations alter the mix of cases treated and influence the frequency, duration, and sequence of clinical activities. Predictive models developed on a given population can rapidly decline in accuracy if patient characteristics change over time, generating data drift that requires continuous updates. Taken together, these challenges demonstrate that data and models in the healthcare domain cannot be considered stable or fixed once and for all.

The dynamic nature of clinical processes, combined with data complexity and human and organizational variability, requires tools capable of supporting not only accurate analysis but also continuous maintenance processes. It is precisely the emergence of these limitations that motivates the need for a methodological framework for the management, repair, optimization, and constant updating of process mining models in healthcare.

1.3.1 Heterogeneous Data Sources

The multiplicity and heterogeneity of available data sources represent one of the main challenges in applying process mining in healthcare. Datasets such as those used in [10, 18] highlight how each data source can differ significantly in terms of recording standards, tem-

poral granularity, structure, and clinical semantics. This variability represents a significant obstacle to the consistent reconstruction of clinical pathways, compromising the quality of process analyses and subsequent predictive applications[35]..

Data integration and homogenization are imperative; the absence of harmonization can induce bias, resulting in misleading or incomplete outcomes during both the process discovery phase and the predictive functionalities of machine learning models. Specifically, in the discovery phase, data heterogeneity can prevent the identification of realistic clinical flows, hindering the extraction of meaningful insights into operational and therapeutic dynamics. From a machine learning perspective, models trained on non-uniform data can develop spurious correlations or overlook clinically relevant patterns, compromising the robustness and reliability[32]. of predictions.

To mitigate these risks, it is necessary to implement semantic mapping strategies, format standardization, and granularity normalization, as well as quality controls aimed at identifying inconsistencies and anomalies across different sources. Only through a systematic process of integration and homogenization can we ensure that clinical process analysis and machine learning applications provide valid, reproducible, and clinically meaningful results.

1.3.2 Incomplete Event Logs

A second significant criticality in the application of process mining and machine learning in healthcare is the incompleteness of event logs. Clinical logs are often characterized by missing or incorrect labels, unrecorded activities, incomplete or inconsistent timestamps, and a general inconsistency in the quality of recordings. These issues regularly emerge in real-world datasets, where the complexity of clinical flows and the distributed nature of healthcare information systems make gaps in event traces inevitable [36].

Addressing these challenges requires robust preprocessing techniques and the integration of advanced algorithms to ensure data quality. By enhancing the reliability of event logs, healthcare organizations can leverage process mining and machine learning more effectively to improve patient outcomes and streamline operations. The types of incompleteness that can be observed are different. One example is the complete absence of certain activities, generating

partial traces that do not faithfully reflect the actual sequence of clinical actions. Second, missing or inconsistently recorded timestamps may compromise the ability to reconstruct the temporal dimension of the process, a crucial element for evaluating care performance and identifying bottlenecks. Finally, incorrect or non-standardized labels introduce further semantic distortions, complicating the correlation between events, resources, and clinical conditions.

The impact of incompleteness on analytical tools is significant. In the process discovery phase, distorted or partial traces can lead to the extraction of process models that are uninterpretable, artifactual, or clinically meaningless. From a machine learning perspective, training on incorrect time series or activity sequences compromises the robustness of models, generating unreliable predictions or those unable to generalize to real-world cases. In particular, the absence of critical information can induce spurious correlations [38] or mask clinically relevant patterns.

To address this structural challenge, the literature has proposed a set of repair and imputation strategies, ranging from rule-based techniques derived from knowledge of the clinical process to statistical or machine learning methods that estimate missing activities or correct temporal inconsistencies. Hybrid approaches, integrating declarative process models, domain knowledge, and machine learning techniques, represent a promising direction for compensating for incompleteness without introducing new biases.

In summary, managing incompleteness in clinical logs is a preliminary technical step and subsequently a fundamental prerequisite for ensuring the reliability of process discovery and the validity of predictions generated by machine learning models in the healthcare field.

1.3.3 Evolving Clinical Protocols

In the healthcare context, clinical processes are inherently dynamic. Care protocols undergo continuous changes due to evolving medical practices, the adoption of new diagnostic and therapeutic technologies, and the adaptation of organizational structures to the needs of patients and staff[18]. This changing nature is directly reflected in event logs, generating variations in the observable behavior of processes and introducing concept drift phenomena that impact

both process mining and machine learning applications.

Several studies, including those on temporal aggregation and temporal feature extraction, show that even relatively small changes in operating procedures can produce significant changes in recorded traces. These findings highlight the importance of continuously updating models and algorithms to accommodate evolving practices. By doing so, organizations can enhance their ability to predict outcomes and improve overall efficiency in patient care and operational workflows. The time windows associated with activities, for example, can compress or expand following organizational changes or workflow optimization, altering the observed temporal distributions. Similarly, the behavior of healthcare professionals or the introduction of decision-support technologies can transform the underlying temporal patterns , with direct repercussions on the sequence, duration, and frequency of clinical activities.

These changes impact the validity and relevance of discovered process models. In the absence of mechanisms capable of detecting and interpreting drift, models become progressively misaligned with current clinical practice, compromising the descriptive and interpretive capacity of process mining. From a predictive perspective, machine learning models trained on outdated historical data risk becoming obsolete, reducing the accuracy of predictions and potentially generating inappropriate clinical recommendations.

Effectively managing these phenomena requires the adoption of adaptive models capable of promptly recognizing drift and updating themselves automatically or semi-automatically. Approaches based on continuous monitoring of temporal distributions, drift detection techniques, and incremental model retraining represent increasingly relevant strategies. In the healthcare context, where variability is structural, these methods play a central role in ensuring that process representations remain consistent with operational practices and that predictive models maintain an adequate level of reliability.

Ultimately, concept drift introduces a need for ongoing maintenance, both for process mining and machine learning models. Only through systematic, evidence-driven updates can we ensure that analyses remain valid, clinically meaningful, and aligned with evolving healthcare protocols.

1.3.4 Continually Changing Patient Populations

The final challenge, relating to the complexity of combining process mining and machine learning in healthcare, is the dynamic nature of the patient population. Unlike relatively stable industrial settings, the pool of patients accessing healthcare facilities constantly fluctuates over time based on demographic, clinical, epidemiological, and seasonal factors. Changes in average age, the prevalence of specific acute or chronic conditions, the severity of cases treated, or the spread of seasonal pathogens directly influence the operational configuration of clinical processes. These variations necessitate adaptable algorithms that can effectively learn from evolving data patterns and provide timely insights. Moreover, integrating these advanced technologies requires a robust framework that ensures data quality and interoperability across diverse healthcare systems.

MIMIC dataset, based on data from emergency departments, clearly illustrates this phenomenon: changes in case composition lead to changes in waiting times, staff workload, and the duration and sequence of clinical activities. For example, when the number of highly complex patients increases, the entire operational flow tends to restructure, expanding the time windows for diagnostic activities, intensifying the need for critical resources, or altering the priorities in performing procedures.

These changes impact time distribution, operating frequencies, and the very structure of process traces, giving rise to data drift and concept drift. Data drift emerges when the distribution of clinical and operational variables shifts over time, while concept drift occurs when the relationship between inputs and clinical outcomes changes as a result of evolving patient profiles or their care needs. As a result, healthcare professionals must continuously adapt their strategies and tools to maintain accuracy in their assessments and interventions. This ongoing adaptation not only requires robust data monitoring systems but also a proactive approach to training and resource allocation to address the dynamic nature of patient care. Both phenomena compromise the validity of models if not actively monitored and managed.

For process mining, these dynamics generate sequences of activities that are no longer consistent with those observed in previous periods, reducing the relevance of reconstructed models and the ability to correctly interpret operational performance. In machine learning,

training on historical populations that no longer reflect current reality can result in a significant loss of predictive accuracy, especially in models dedicated to triage, length of stay prediction, or decision support [18].

Consequently, it is necessary to adopt strategies for continuously updating analytical models and operational workflows. Mechanisms for statistically monitoring changes in the patient population, periodic retraining pipelines, adaptive models, and incremental approaches are essential tools for maintaining alignment between process representation, operational reality, and emerging clinical needs.

In summary, the constant change in patient populations represents a critical dimension of healthcare drift, making a flexible and adaptive approach essential for both process mining and machine learning models to maintain their relevance and reliability over time.

1.4 Why Data and Models Require Continuous Maintenance

The studies analysed converge in highlighting a central and inescapable principle in the healthcare domain: neither data nor analytical models can be considered static entities. On the contrary, they are intrinsically subject to continuous transformation, driven by evolving clinical practices, changing patient populations, updated organizational protocols, and the introduction of new technologies. This dynamic nature makes constant maintenance essential, aimed at ensuring that the analyses produced remain current, reliable, and consistent with the operational reality in which they are applied. Methodological inertia, in the absence of such updates, progressively leads to distorted interpretations, obsolete models, and potentially erroneous decisions.

The need for maintenance arises from a variety of factors that influence the quality and effectiveness of process logs and predictive models over time:

Semantic Degradation (Log and Process Mining)

The repair phase of missing activities demonstrates how errors, biases, or inconsistencies in operational logs compromise the integrity of the data and, conversely, the validity of the

entire analytical process. Without constant monitoring of log quality, consisting of semantic checks, reconciliation, and corrections guided by process knowledge, the correspondence between actual clinical behaviour and recorded traces tends to deteriorate. This degradation compromises both the descriptive capacity of process mining and the reliability of the models derived from it.

Structural Degradation (Model Complexity)

Evidence derived from sampling and aggregation techniques shows how the evolution of clinical processes can lead, over time, to an increase in the structural complexity of process models. Models that are not periodically reviewed risk becoming excessively branched, difficult to interpret, and costly to maintain. The lack of systematic structural control leads to a progressive loss of readability, manageability, and explanatory capacity, with repercussions on the understanding of clinical pathways and audit activities.

Predictive Degradation (Feature Drift and Concept Drift)

Analyses of temporal features demonstrate how the predictive relevance of variables is not stable over time. The emergence of new clinical patterns, the evolution of caseloads, or variations in the temporal distribution of activities generate feature drift and concept drift. A predictive model trained in a past context does not necessarily reflect the current behavior of the system and, in the absence of regular updates, experiences a progressive decline in performance. The impact can be significant: reduced accuracy, increased classification errors, and the risk of generating inappropriate clinical recommendations.

To mitigate these risks, it is necessary to adopt a perspective according to which the entire Process Mining (PM) and Machine Learning (ML) cycle is, by its nature, iterative and inconclusive. Model updating is not an extraordinary operation but a structural component of the analytical lifecycle. Specifically, predictive models must undergo periodic retraining based on two main conditions:

1. The availability of new clinical data that more faithfully reflects current operations.

2. The detection of a performance decline below defined thresholds, indicating the presence of drift or structural changes in the population or process.

Adopting this continuous cycle with observation, discovery, evaluation, and update phases requires developing an organizational and technical infrastructure that ensures:

- Auditability : The ability to track and explain every change made to data, models, and workflows, making the evolution of the analytical system transparent.
- Continuity : The implementation of automated pipelines for performance monitoring, drift identification, model retraining, and deployment, minimizing manual intervention and reducing response times.
- Governance : The definition of formal frameworks that establish roles, responsibilities, audit frequency, performance alert thresholds, update criteria, and rollback procedures. Governance of the data-model cycle thus becomes an essential element for ensuring security, with quality and reliability.

The quality and reliability of clinical decision support systems depend on a constant commitment to maintenance and continuous updating. In a highly dynamic environment like healthcare, machine learning is not a static solution but rather an evolutionary path.

1.5 Motivation and Introduction of the Proposed Framework

The analysis presented in the previous sections highlights a fundamental characteristic of healthcare processes and data: they are inherently dynamic and continuously evolving. Clinical workflows change as a result of updated medical guidelines, organizational reconfigurations, technological innovations, and fluctuations in patient populations. At the same time, the data generated by healthcare information systems are often heterogeneous, incomplete, noisy, and subject to temporal and semantic drift. In such a context, analytical models based on process mining and machine learning cannot be conceived as static artifacts, but must instead be treated as living entities that require systematic monitoring, adaptation, and maintenance.

The primary motivation for this work arises from the observation that many existing approaches in healthcare process mining focus on isolated phases of the analytical pipeline—such

as event log preprocessing, model discovery, or predictive analysis—without explicitly addressing their long-term sustainability. While these contributions provide valuable solutions to specific problems, they often lack an integrated perspective capable of managing the full lifecycle of data and models in the presence of continuous change. As a consequence, models that are initially accurate and clinically meaningful may progressively lose validity, becoming misaligned with real operational practices and potentially leading to misleading interpretations or unreliable predictions.

The need for a structured and holistic framework is further supported by empirical evidence observed in healthcare process analytics. Log repair techniques show that semantic inaccuracies and missing activities can severely compromise downstream analyses if not addressed in a systematic manner. Sampling and aggregation strategies highlight how uncontrolled process complexity and temporal noise hinder both interpretability and scalability. Studies on temporal features and predictive monitoring emphasize that feature relevance and model performance are not stable over time, but are strongly influenced by changes in execution patterns and contextual conditions. Finally, real-world analyses based on large clinical datasets confirm that the integration of process mining and machine learning is feasible and beneficial only when supported by robust data management practices and continuous update mechanisms.

These observations motivate the introduction of a unified framework explicitly designed to support the continuous maintenance of healthcare process analytics. The proposed framework aims to coordinate data acquisition, standardization, quality assessment, and repair with model discovery, optimization, and predictive analysis, embedding all these activities within a governance-oriented lifecycle. Rather than treating maintenance as an ad hoc or reactive activity, the framework promotes a proactive and cyclical approach in which data and models are periodically evaluated, validated, and updated according to predefined criteria and performance indicators.

In particular, the framework is designed to address three key objectives. First, it seeks to preserve data reliability by systematically managing heterogeneity, incompleteness, and noise in clinical event logs. Second, it aims to control model complexity and relevance by supporting adaptive discovery and optimization strategies that remain interpretable for

clinical stakeholders. Third, it enables the continuous alignment of predictive models with evolving clinical practices and patient populations, mitigating the effects of concept drift and performance degradation.

This theoretical and practical limitation in current literature defines the core research gap addressed by this thesis: the absence of a standardized, governance-aligned methodology for the longitudinal maintenance of process-oriented clinical analytics. To bridge this gap, the research investigates how heterogeneous clinical data can be effectively acquired and standardized for process mining (RQ1) and how data quality can be ensured in clinical event logs (RQ2), furthermore, the study explores how complex healthcare processes can be represented in an interpretable and scalable way (RQ3), how predictive models can be trained and maintained to anticipate clinical outcomes (RQ4), and finally, how process mining systems can remain trustworthy, auditable, and aligned with clinical governance principles (RQ5).

By integrating these objectives into a single methodological structure, the proposed framework provides a coherent foundation for sustainable process mining and machine learning applications in healthcare. It supports not only descriptive and diagnostic analyses, but also predictive and adaptive functionalities that can evolve alongside the clinical environment. In doing so, this work contributes a systematic approach to ensuring that healthcare analytics remain accurate, trustworthy, and operationally relevant over time.

CHAPTER

2 State of the Art

In recent years, the application of process mining in healthcare has garnered increasing attention. Its adoption is driven by increasing digitalization and the availability of large volumes of clinical data, which enable hospitals, clinics, and other healthcare providers to optimize workflows, lower operational costs, and improve resource management. The value of process mining lies in its ability to detect inefficiencies in processes, foster the standardization of clinical practices, support evidence-based decision-making, and ultimately enhance the overall quality of care delivered.

This chapter provides an overview of the state of the art in process mining applied to healthcare; the aim is to explore how these techniques are currently employed, the benefits they can generate, and the challenges that researchers and practitioners still face. Recent systematic mappings and updated perspectives have highlighted the importance of mapping the patient's journey to improve clinical outcomes [6, 20, 23]. The literature survey conducted in this study identifies the main thematic areas that reflect the current state of research on process mining in healthcare and outline its future perspectives.

To address these aspects, the chapter is structured into five principal sections:

- **Research Method:** describing the approach adopted for the literature review, including the search strategy, inclusion and exclusion criteria, and the procedure followed for the classification and analysis of the selected studies.

- **Process Mining for the Optimization of Clinical Workflows:** with particular em-

phasis on the role of process mining in optimizing clinical workflows and identifying inefficiencies.

- **Algorithms and Techniques Most Commonly Used in Healthcare:** focusing on the level of adoption and the effectiveness of different process mining algorithms and techniques applied in specific healthcare contexts.

- **Clinical Data and Heterogeneity of Formats:** addressing the types and characteristics of clinical data used in studies, with particular attention to the issues associated with their heterogeneity and complexity.

- **Open Challenges and Research Perspectives:** highlighting the main critical issues related to data management and the application of process mining in healthcare, and outlining the open challenges and possible future research directions.

2.1 Research Methodology

This chapter outlines the current scientific landscape of process mining in healthcare, highlighting both the progress made and the challenges that remain, such as data quality, process complexity, and system integration. To conduct the state-of-the-art review, a systematic methodology was adopted, focusing on key research areas, the selection of relevant databases, and rigorous filtering procedures, following the principles of Kitchenham (2004).

In this context, three thematic areas were first identified to structure the analysis and guide the study, each addressing a specific aspect of process mining in healthcare.

The first theme addressed concerns the identification of the main research topics emerging from primary studies on process mining in healthcare. Mapping the topics covered in the literature makes it possible to understand which aspects have attracted the greatest attention from scholars. This is particularly important in the healthcare domain, where processes are complex and multidimensional and therefore require detailed analysis to improve efficiency and quality of care. Highlighting the most prevalent themes allows emphasis on areas of major impact, such as the optimization of hospital workflows, the reduction of waiting times, the

improvement of patient management, and the analysis of variability in treatment pathways, while at the same time revealing less explored areas that may represent promising directions for future investigation.

The second focus concerns the examination of the different approaches and algorithms applied within the healthcare domain, which is fundamental to assessing the quality and robustness of the available evidence. Establishing this foundation is essential to ensure that conclusions are grounded in reliable data and valid methodologies, thereby preventing the adoption of practices based on insufficient or unconvincing evidence. A systematic identification and classification of process mining algorithms employed in diverse healthcare contexts makes it possible to evaluate the breadth and heterogeneity of the techniques in use, as well as to highlight the prevalence and distribution of specific applications.

Process mining algorithms can serve multiple purposes within healthcare, including the identification of patterns in clinical data, the detection of bottlenecks along care pathways, and the suggestion of targeted interventions to improve operational efficiency. Examining the distribution of these algorithms makes it possible to determine which methodologies prove most effective in specific healthcare contexts. At the same time, such an analysis helps to highlight the inherent limitations of existing approaches and to identify areas where further methodological development is required.

This line of inquiry is particularly relevant in the healthcare domain, which is characterized by substantial variability across organizational and clinical contexts. Moreover, the sector's constant evolution demands continuous innovation to address emerging challenges and adapt process mining techniques to new conditions. Providing a systematic perspective that connects specific process mining techniques with healthcare domains can therefore offer practical value for both researchers and practitioners, while also supporting the development of a best practice framework to guide future studies and inform applied initiatives in the sector.

The third focus concerns the identification of the types of data employed in process mining within healthcare, which is essential for understanding the variety and complexity of the information available. Relevant data sources may include electronic patient records, laboratory results, clinical workflows, treatment information, and other forms of operational data. Each

category of data possesses distinct characteristics that influence the ways in which it can be analyzed and interpreted using process mining techniques.

Comprehending the attributes of healthcare data—such as structure, granularity, quality, provenance, completeness, and privacy management—is crucial for selecting appropriate process mining algorithms and for designing studies that produce valid and reliable results. A thorough understanding of these aspects allows researchers and practitioners to fully exploit the potential of process mining, enhancing healthcare processes and generating meaningful insights.

After defining the research directions, the method to be used and the sources from which to extract the studies to be included in the analysis were established.

The process was structured into three crucial phases, as illustrated in Figure 2.1.

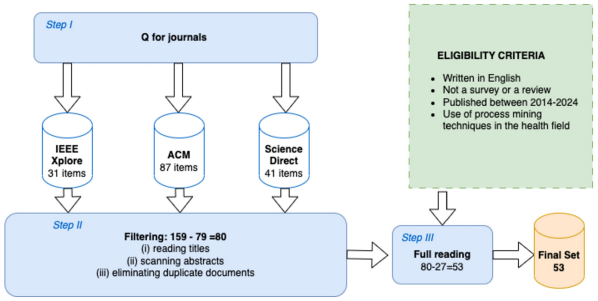

Figure 2.1: Approach used for construction of state of art.

In the first phase, specific queries were executed on the selected databases to collect relevant documents.

The use of targeted queries was crucial to identify and select the most relevant studies for the state of the art. The queries were carefully constructed, using keywords extracted from the previously identified research themes, in order to capture with maximum accuracy the works most consistent with the topics of interest. The choice of keywords and their combination was designed to optimize sensitivity and specificity.

Querying databases with specifically designed queries makes it possible to filter a large number of publications and focus on those that best fit the research areas. This approach helps reduce the risk of omitting important studies and ensures that the study is based on a comprehensive collection of relevant evidence. Before choosing the sources to be used, in

order to have a complete yet essential vision, several reference databases were selected.

These databases were selected for their breadth of coverage in the field of process mining and their ability to host a wide range of scientific publications in the healthcare domain. Below are the databases queried:

- IEEE Xplore: A digital library provided by the Institute of Electrical and Electronics Engineers (IEEE), one of the world's largest professional organizations for the advancement of technology. Includes journal articles, conference proceedings, technical standards, ebooks, and educational courses. Its main topics include Electrical engineering, electronics, information technology, telecommunications, robotics, energy, biomedicine, and other related areas. Continuous updates with new publications and technical documents.

- ACM Digital Library: a comprehensive collection of scientific and technical literature produced by the Association for Computing Machinery (ACM). It mainly covers the field of computer science and related disciplines such as artificial intelligence, software engineering, computer networks, cybersecurity, human-computer interaction, and much more. Includes journals, conference proceedings, technical journals, newsletters, and books published by the ACM.

- ScienceDirect: a leading academic research platform operated by Elsevier, a leading global publisher of scientific, technical, and medical literature. It hosts scientific journal articles, book chapters, and other academic publications from Elsevier. It covers a wide range of scientific and technical disciplines, including life sciences, physical sciences, health sciences, engineering, social sciences, and humanities. It offers powerful search and navigation tools, with the ability to filter results by document type, publication date, author, and more. Additionally, it includes features for citation management and access to impact metrics.

Another important parameter that influenced the investigation was the careful choice of the time frame to be assessed. A 10-year time horizon was chosen because it offers a balance between breadth and relevance, allowing the inclusion of a sufficient number of studies to

obtain a comprehensive view without making the data obsolete. This time interval is long enough to include significant developments in the field but not so long as to make it difficult to interpret current trends. This first research phase produced a total of 159 articles, of which 87 came from ACM, 31 from IEEE Xplore, and 41 from ScienceDirect.

After the first selection phase, a filtering process was carried out, applying rigorous inclusion criteria (IC) and exclusion criteria (EC) to filter the initially selected documents, as detailed in Table 2.1.

Criterium ID	Description
4*Inclusion Criteria	
IC1	Studies written in English
IC2	Studies published in the range 2014–2024
IC3	Studies published in a press or in a journal
IC4	Studies should use process mining techniques in the healthcare domain
3*Exclusion Criteria	
EC1	The studies are bibliographic surveys or systematic reviews
EC2	The studies do not involve Process Mining techniques
EC3	The studies do not concern the healthcare domain

Table 2.1: Inclusion and Exclusion criteria

The criteria used are divided into inclusion criteria and exclusion criteria. The first ensured that the selection included only articles written in English (IC1), within the period 2014/2024 (IC2), published in journals (IC3), and using process mining techniques in the healthcare sector (IC4). The exclusion criteria, instead, indicated the rules adopted: exclusion of studies that were already reviews or surveys (EC1), exclusion of studies that did not use process mining techniques (EC2), and exclusion of studies not related to the healthcare domain (EC3).

This approach allowed us to maintain the focus on studies that directly contribute to the analysis and optimization of healthcare processes, ensuring relevance and applicability. The result of this second filtering phase involved (i) reading the titles, (ii) scanning the abstracts, and (iii) eliminating duplicate documents, leading to a total of 80 articles.

Finally, in the third phase, a detailed analysis of the studies included in the review was carried out, in which 53 documents were considered because they met the eligibility criteria.

After selecting the studies relevant to the analysis, the data extraction phase followed. For this purpose, a standardized protocol was used that included labels such as the study objective,

the process mining methods used, the types of healthcare data analyzed, and the main results. Each article was examined and the necessary data were extracted to answer the previously identified research questions. The extracted data were then organized in a standardized format, with each column representing a specific characteristic identified in the individual studies. After data extraction, a final review was conducted to resolve any discrepancies, thus ensuring an objective and rigorous approach. This guaranteed the consistency of the extracted data and facilitated the final synthesis.

The results of the studies were then categorized and compared to identify common trends, methodological differences, and potential areas of innovation. Therefore, in summary, after the initial identification of the articles, a multistep selection process was followed to select the relevant studies. First, duplicates were eliminated and then titles and abstracts were screened to assess compliance with the inclusion criteria. Subsequently, a full-text review of the selected articles was conducted to confirm relevance, thereby ensuring an objective and rigorous approach.

From this initial phase, initial results are already emerging, making the state of the art more precise and concrete. Figure 2.2 illustrates the temporal distribution of the studies included in the analysis. This graph allows us to track trends and variations in the number of studies published over the period considered. Specifically, it allows us to identify significant temporal trends, such as peaks or declines in research activity, periods of increased interest or attention to specific aspects of the subject, and any interruptions or anomalies in scientific production. Analyzing this data allows us to better understand the evolution of the field and the challenges faced over the years. In the figure, the years are represented on the x-axis, while the number of contributions is represented on the y-axis, indicating both the total number of articles in blue and the percentage in yellow. The analysis shows a significant increase in the number of relevant articles every two years between 2018 and 2020, stabilizing in 2022/2023, and decreasing in the final year considered.

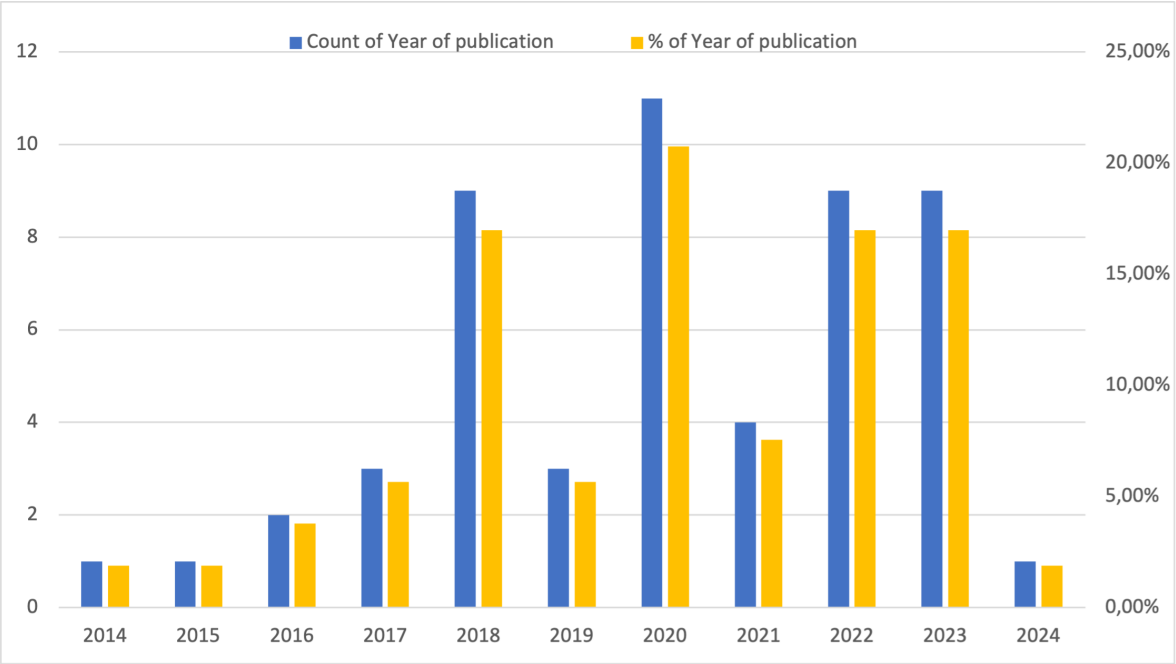

Figure 2.2: Yearly Analysis of Paper Distribution.

In addition, a graph was created on the distribution of articles by publisher, Figure 2.3 to understand the impact and influence of different publishers in the field under consideration. This graph shows how many publications have been produced by each publisher, thus providing an indication of editorial leadership in the sector. Furthermore, analyzing the distribution of articles by publisher can help identify editorial trends, author preferences, and the reputation associated with each publisher in the specific research field. These data are essential to evaluate the diversity and quality of available information sources and to guide strategic decisions on the dissemination and access to scientific knowledge. Therefore, Figure 2.3 presents the distribution of the included articles by publisher, where the publishers' names are shown on the x-axis and the counts are represented on the y-axis. The blue bars indicate the absolute number of articles per publisher, while the yellow bars express the relative percentage. As can be seen in the graph, the figure highlights the prevalence of articles published by Elsevier within the research topic, followed by ACM and IEEE in equal measure. The contribution of the International Society for Pharmacoeconomics and Outcomes Research and The Lancet Discovery Science is limited.

Finally, Figure 2.4 shows the distribution of articles according to the countries of the

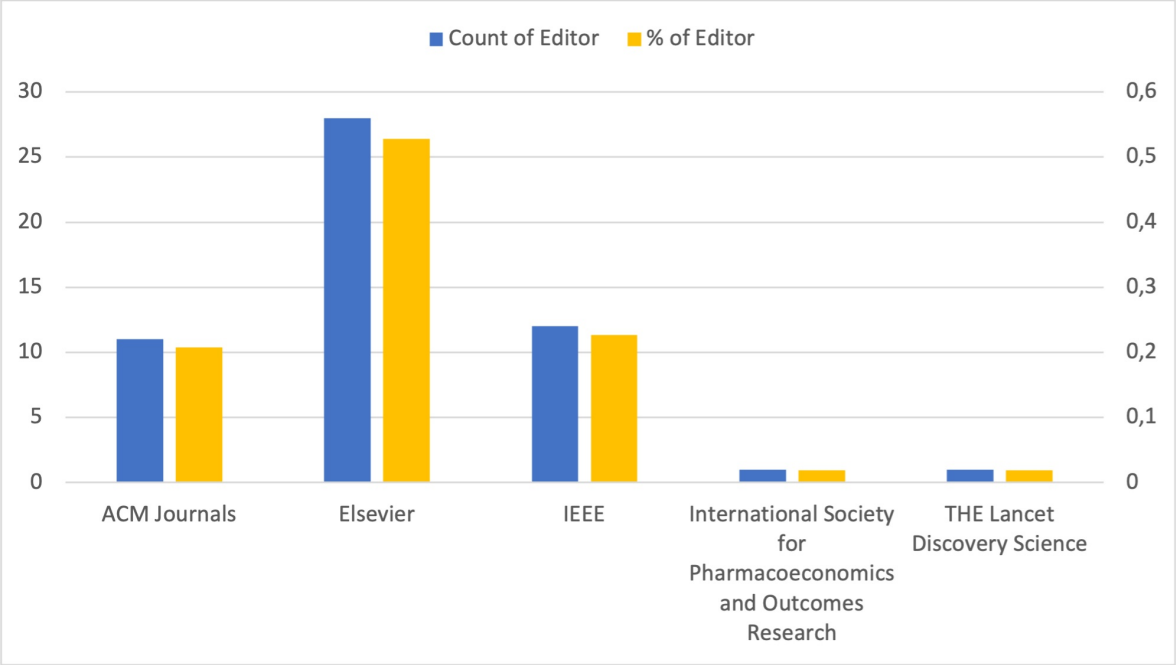

Figure 2.3: Distribution of Papers by Publisher.

authors involved in the study. Each bar represents a specific country, while the y axis indicates the number of articles published by authors in that country. This graph is fundamental as it allows us to understand the geographical origin of the authors of the articles included in the study. Analyzing the distribution of articles by country provides information on the geographical diversity of publications, highlighting the regions of the world where greater research activity in this specific field is recorded. This information can be useful for identifying regional trends, comparing research activity across countries, and assessing the geographical impact of the topics addressed in the analyzed articles. From the figure, it is clear that the countries producing the largest number of process mining studies in the healthcare domain are Germany and the United States, followed by Canada, the United Kingdom, South Korea, and China.

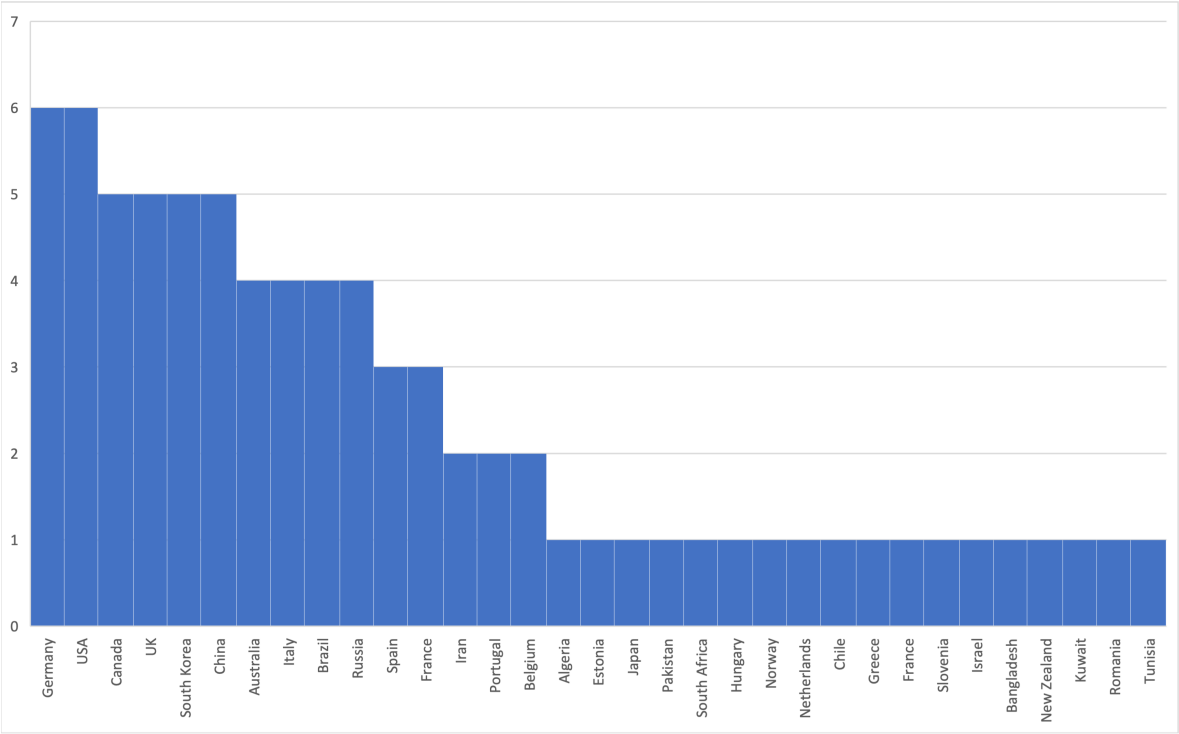

Figure 2.4: Country-wise Distribution of Papers by Authors.

2.2 Overview of Process Mining Areas for Healthcare Process

Process mining is widely regarded as an effective approach for enhancing healthcare processes by integrating data-driven analysis with clinical workflow optimization. To elucidate the application of this methodology in healthcare, a quantitative analysis was performed focusing on the principal areas of process mining. Based on the classification outlined in [9], four primary areas were examined: process discovery, conformance checking, process enhancement, and supporting areas. This enables a more precise understanding of research trends and priorities within the field. By analysing the distribution of publications across these categories, the study try to determine which aspects of process mining have attracted the most attention in healthcare and how these areas contribute to the overarching objective of optimizing clinical workflows.

A quantitative assessment was conducted to identify the most frequently explored process

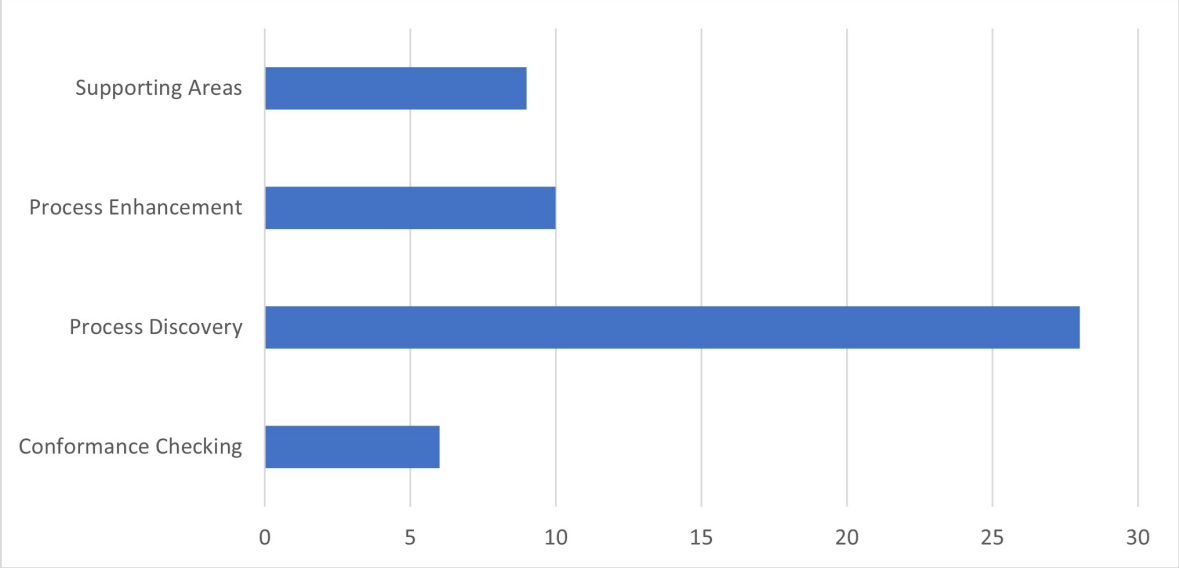

Figure 2.5: Number of papers for each Process Mining topic applied in healthcare.

mining topics in healthcare by counting the number of documents associated with each area. The analysis focused on the four categories defined in [9]: process discovery, conformance checking, compliance, enhancement, and support. The support area includes papers that do not directly describe one of the three main types of process mining,but explore other related questions to the implementation of process mining projects.

Figure 2.5 presents the results, with a bar chart displaying the four topics on the y-axis and their corresponding publication counts on the x-axis. The analysis indicates that Process Discovery is the most frequently investigated topic, with nearly 30 publications, reflecting substantial interest in deriving process models from data. Process Enhancement and Supporting Areas have a moderate number of studies, suggesting that while process improvement and support-related topics are relevant, they are less central than process discovery. Conformance Checking is the least represented, with fewer than ten studies, indicating that research on verifying process compliance or models remains relatively limited.

Figure 2.6 illustrates the annual distribution of articles by topic from 2014 to 2024, using a stacked histogram. Each color represents a specific research category: yellow for Supporting Areas, gray for Process Enhancement , orange for Process Discovery, and blue for Conformance Checking. Between 2014 and 2016, the number of publications was very

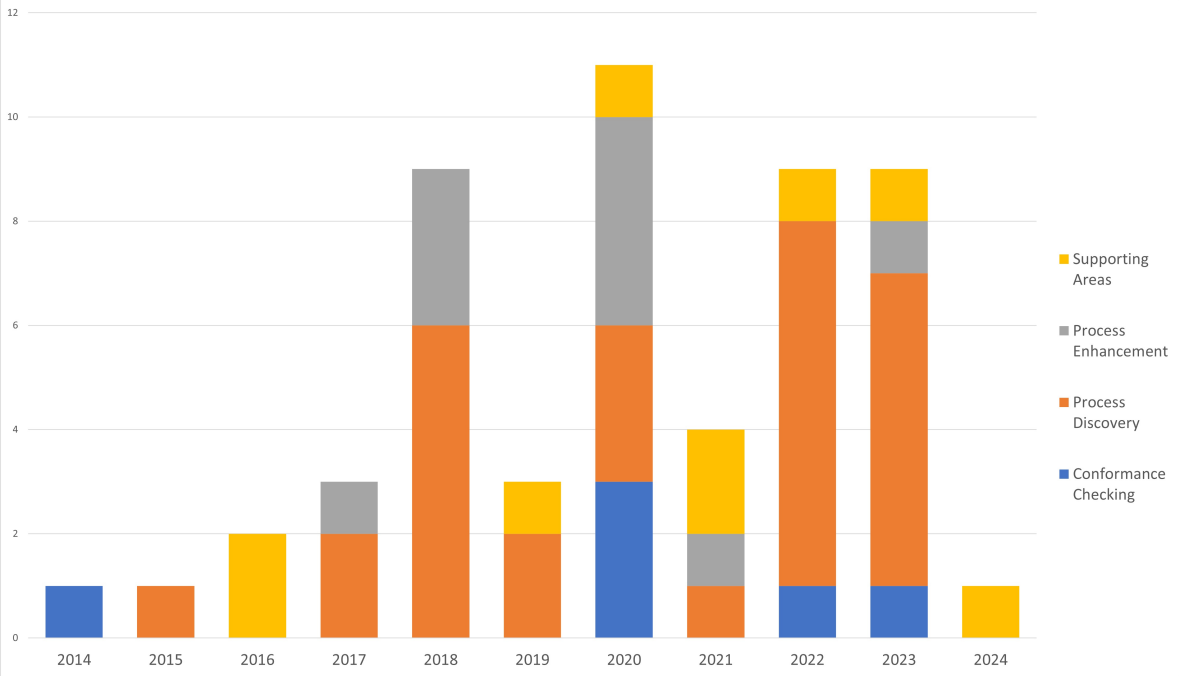

Figure 2.6: Annual distribution of papers per topic.

low, averaging one or two per year, mainly focused on support-related studies. A gradual increase occurred in 2017, accompanied by diversification into new topics such as process enhancement and process discovery. The year 2018 marks a significant peak in publication activity, particularly in process enhanecement and process discovery . In 2020, the number of publications reached an all-time high, with strong growth across all areas—especially in conformance and enhancment. The years 2022 and 2023 show a recovery in publication rates, with balanced representation across all topics and a slight predominance in support and discovery areas. Overall, the figure indicates a general upward trend in research activity until 2020, followed by fluctuations in subsequent years, yet maintaining steady interest across all major areas.

Figure 2.7 provides a clear and insightful overview of current research priorities in Process Mining, organizing the examined articles into four foundational macro-categories, each encompassing specific areas of focus. The Conformance Checking area, while stable but not predominant, concentrates on assessing how closely real-world process executions align with their intended models. Specific topics within this category include general conformance checks and the analysis of process quality and performance metrics. Moving toward the evolution

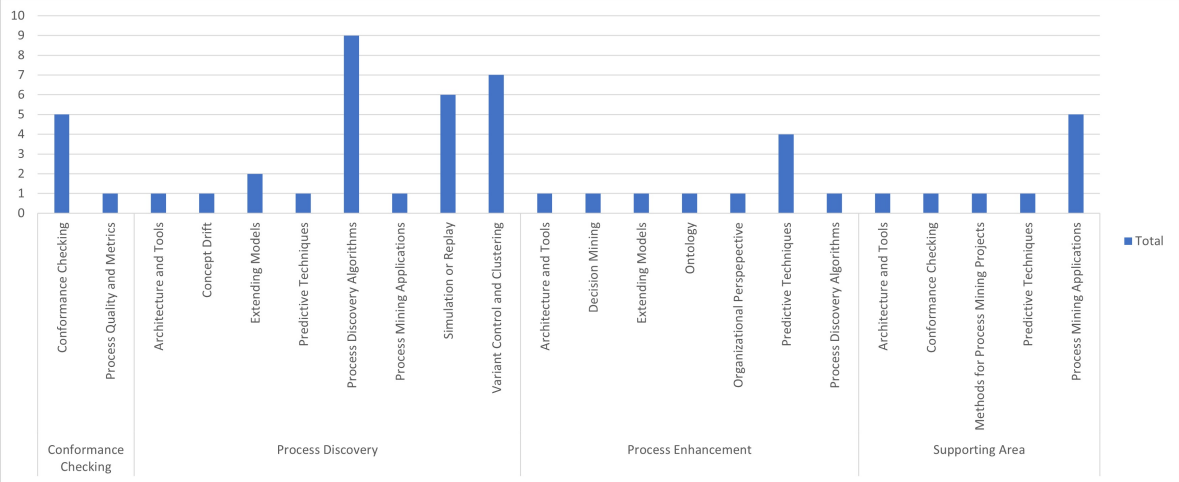

Figure 2.7: Distribution of specific topics by main topic.

of process execution, the Process Enhancement category aims to actively improve existing workflows. The articles in this area primarily focus on Predictive Techniques, which represent the most prominent topic, alongside more conceptual studies on Ontology, the Organizational Perspective of process improvement, and the study of Concept Drift, which investigates how process behaviors evolve over time. The Supporting Area encompasses the transversal themes necessary for the implementation of process mining in practice. It mainly includes works on Comprehensive Process Mining Applications and Methods for Process Mining Projects, reflecting the methodological and applied dimensions of the field. However, the core of research activity clearly resides in the Process Discovery macro-category. This area is by far the most extensively studied and productive, characterized by a dominant focus on Process Discovery Algorithms. The prominence of this topic demonstrates the research community's strong commitment to developing more accurate, robust, and intelligent tools capable of automatically extracting reliable process models from complex event logs. Complementing this primary effort are key topics such as Simulation or Replay and Variant Control and Clustering, which emphasize a sustained interest in understanding process deviations and predicting future process behavior.

The temporal distribution of publications by topic from 2014 onward was also assessed (see Figure 2.8). Each sub-figure represents one of the four process mining areas: (a) Conformance Checking, (b) Process Discovery, (c) Process Enhancement, and (d) Supporting Areas. The

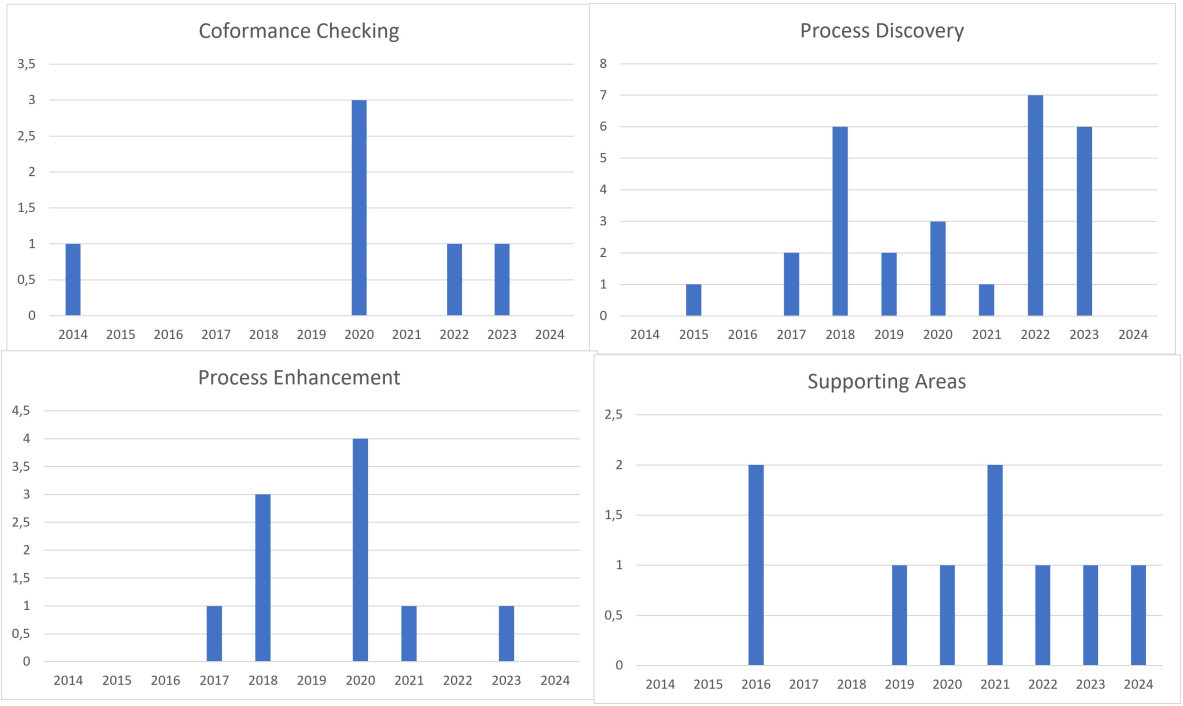

Figure 2.8: Temporal distribution of documents by topic.

Conformance Checking sub-figure shows a notable peak in 2020, exceeding 3.5 publications, while other years such as 2014, 2021, and 2023 display lower values, indicating intermittent attention with a surge in 2020. The Process Discovery sub-figure reveals a continuous upward trend from 2016 to 2021, with peaks in 2019 and 2021, each reaching nearly eight publications. Interest in this area remained high through 2023 and 2024, confirming its ongoing relevance. For Process Enhancement, the primary focus occurred between 2017 and 2020, with peaks in 2018 and 2020, both exceeding four publications, followed by a decline in 2021 and a modest resurgence in 2023 and 2024. The Supporting Areas sub-figure displays a more irregular pattern, with an initial peak in 2015, a decline, and renewed growth in 2019 and 2021. From 2021 to 2024, publication levels stabilized at approximately 1.5 to 2 papers per year, indicating steady but not increasing interest. These trends illustrate distinct temporal dynamics across the four research domains and highlight shifts in research priorities over time.

To complement the analysis , the study also examined whether certain disease types were more frequently investigated within the process mining literature. Figure 2.9 displays a bar chart showing both the absolute number (blue bars) and the percentage (yellow bars) of

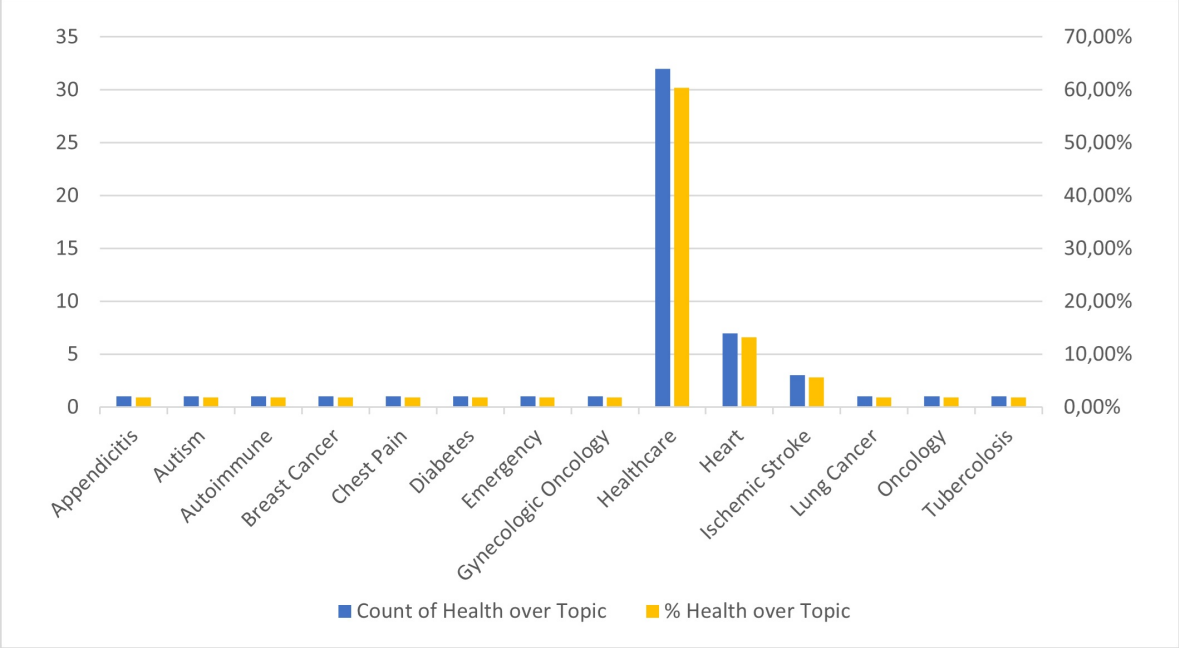

Figure 2.9: Distribution of documents by specific medical field.

publications related to specific diseases. The x-axis represents the various disease categories, while the left and right y-axes indicate, respectively, the total number and the percentage of articles for each condition. The results show that, for most diseases, the number of publications is relatively balanced. However, notable exceptions exist. The "Healthcare" category, encompassing studies that address general healthcare processes, has a significantly higher number of publications than disease-specific topics. This finding suggests a predominant research interest in generic healthcare workflows, likely due to their broad applicability. Heart disease and ischemic stroke also stand out with higher publication counts, possibly reflecting their prevalence and clinical importance. Conversely, other medical fields show relatively uniform research output, indicating balanced scientific attention across multiple conditions. In summary, the analysis highlights not only the overall distribution of research across disease categories but also specific areas where academic interest has been particularly concentrated. While studies on general healthcare processes, heart disease, and ischemic stroke appear most prolific, other disease areas continue to receive meaningful attention, ensuring comprehensive coverage within the scientific literature.

2.3 Application of Process Mining Algorithms in Healthcare

Understanding which process mining algorithms have been proposed and the extent to which they have been applied in specific healthcare contexts provides valuable insight into the evolution of the field. Such analysis helps identify emerging research trends and recognize well-established techniques. Moreover, it highlights both the areas that have been extensively explored and those that present opportunities for future innovation. This perspective is particularly useful for practitioners aiming to implement process mining solutions within healthcare organizations, as it supports the selection of algorithms most suitable for their specific operational and clinical needs. Figure 2.10 effectively highlights the prevailing methodological approach within Process Mining research by comparing the usage of existing techniques versus the proposal of custom or novel techniques. The graph reveals a pronounced trend: the vast majority of articles, approximately 40, rely on methodologies already established in the literature. In contrast, only a smaller segment, around 10 articles, proposes entirely new techniques. This significant imbalance suggests that while there is an appetite for innovation in the field, current research is primarily centered on the application, adaptation, and validation of known methods rather than fundamental algorithmic discovery.

To support this observation, a deeper evaluation of the specific techniques employed confirms that the field is far from stagnant. We observe a notable diversity of algorithms in use, underscoring the dynamic and evolutionary nature of Process Mining, where various methodologies are explored and applied across heterogeneous contexts. More specifically, the algorithms enjoying the highest popularity and frequency of use are those based on trees (such as Inductive Miner or decision trees), Support Vector Machines (SVM), the Fuzzy Miner (often utilized for simplified visualization), and K-means (commonly employed for process variant clustering).

In summary, the research demonstrates a balanced yet grounded approach: there is a strong methodological reliance on existing techniques for robust validation and practical application, coupled with a varied and flexible use of these tools, indicating a mature field that is actively exploring the full potential of its established foundations.

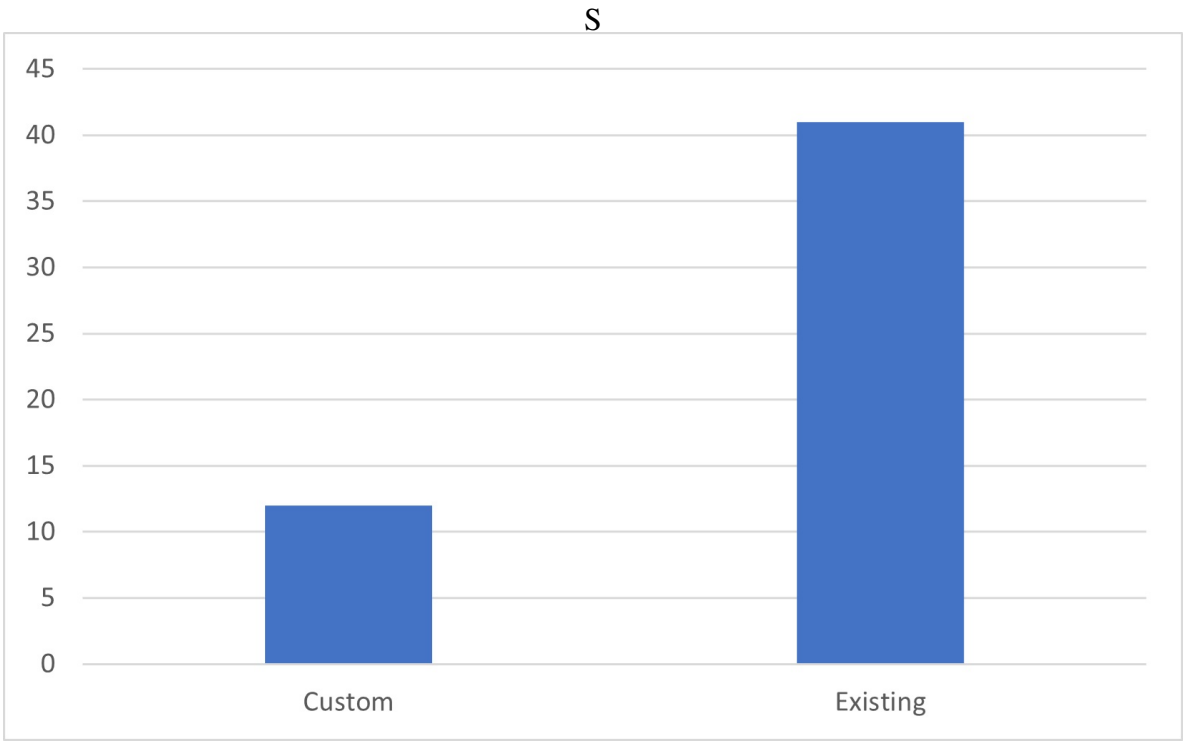

Figure 2.10: Number of articles using custom techniques or already known techniques.

Finally, Figure 2.11 illustrates the distribution of algorithms across different application domains, highlighting which ones are most popular and widely adopted. This visualization is particularly valuable for understanding the preferences of the research community, as it provides insight into which algorithms are perceived as more effective or intuitive for specific healthcare applications.

The analysis reveals a clear predominance of algorithms developed and applied within the general healthcare domain, followed—though to a lesser extent—by the heart disease and diabetes fields. This finding suggests that the broader healthcare domain has attracted the greatest research attention, likely due to the significant practical impact that process mining can have on optimizing both clinical and managerial workflows. Such interest underscores the potential of process analysis techniques to enhance efficiency, reduce waiting times, and ultimately improve patient outcomes.

Among the various algorithmic categories, Predictive Techniques and Process Mining Applications emerge as the most frequently used. Predictive Techniques are particularly relevant in healthcare, as they enable the anticipation of critical events, improve diagnostic

accuracy, and support the personalization of treatments based on patient risk profiles. Conversely, Process Mining Applications provide a detailed understanding of workflows, helping to identify bottlenecks and inefficiencies within clinical processes.

The predominance of these two categories suggests that, despite the availability of numerous alternative algorithms, the research community tends to favor those with demonstrated effectiveness, ease of implementation, and greater interpretability—key attributes in healthcare contexts, where decision-making must be supported by transparent and comprehensible evidence. Additionally, the presence of algorithms in specialized domains such as heart disease and diabetes likely reflects specific clinical needs, where process analysis can facilitate the monitoring of chronic conditions and the identification of optimal care pathways.

2.4 Clinical Data Types and Format Heterogeneity

Understanding the types of data used in process mining analyses is fundamental for evaluating the quality, accuracy, and relevance of the results. In process mining, data serve as the foundation upon which models are discovered, conformance is checked, and workflows are enhanced. Therefore, a thorough comprehension of their nature, structure, and origin directly affects the validity and interpretability of the analyses. In healthcare, clinical data are particularly diverse and heterogeneous, encompassing electronic health records, laboratory results, sensor-generated signals, administrative data, and more. Each type of data carries specific characteristics—such as granularity, completeness, timestamp precision, and format—which can significantly influence the performance of process discovery algorithms, the reliability of conformance checking, and the effectiveness of process improvement initiatives. Recognizing and properly handling this heterogeneity is thus a critical step in ensuring that process mining yields meaningful, accurate, and actionable insights for healthcare applications.

2.5 Summary and Research Gaps

To conclude the state of the art analysis, this section provides a critical synthesis of the current literature . From a systematic review of contributions published between 2014 and 2024 , it

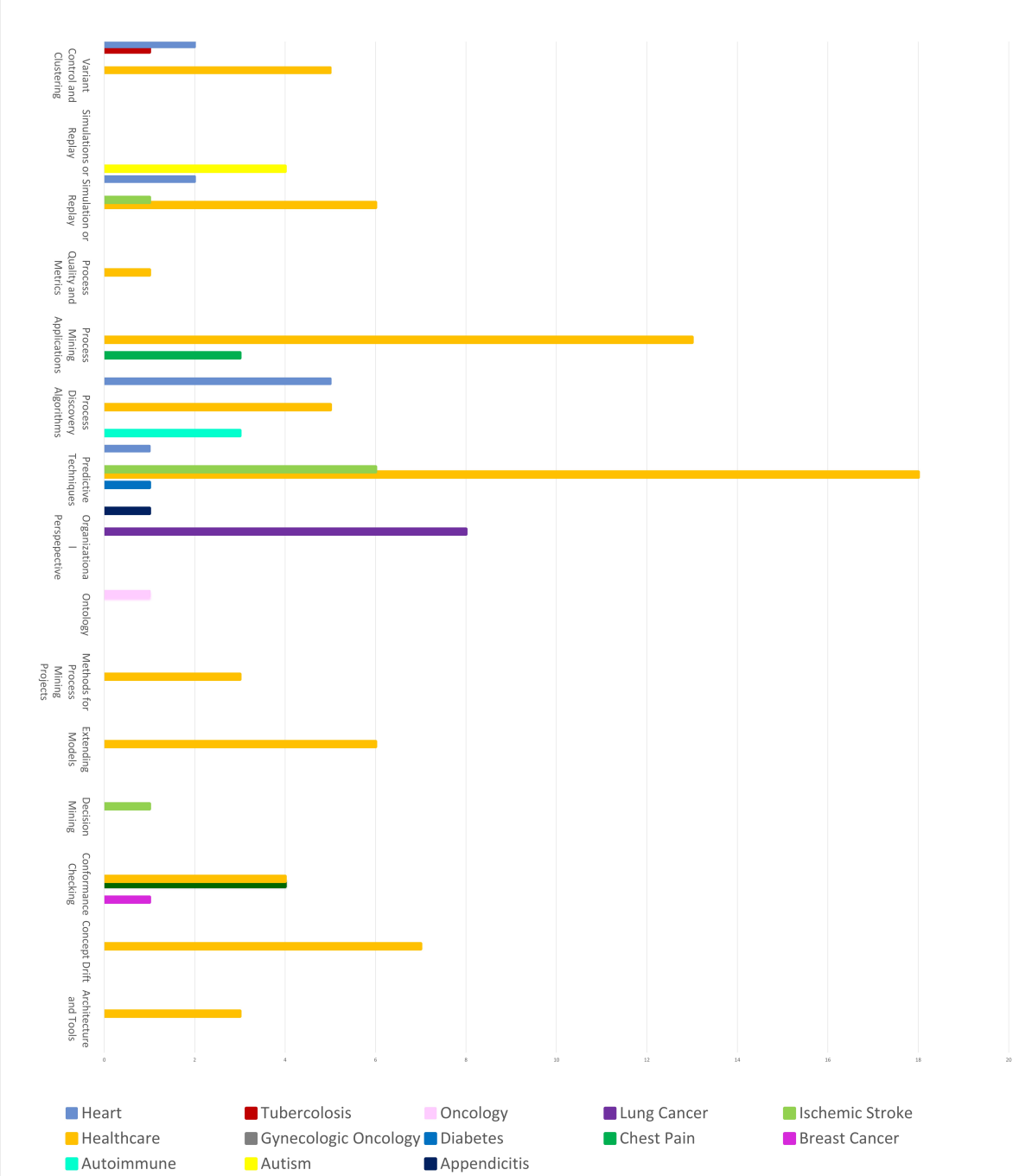

Figure 2.11: Distribution of algorithms by specific topic health.

emerges that , although the field of Healthcare Process Mining has matured significantly, most research remains focused on the initial process discovery phase or the prediction of isolated clinical outcomes , therefore, there remains a significant shortage of models that explicitly address the problem of concept drift in complex hospital environments. Current approaches often fail to provide maintenance tools that are interpretable by clinicians and computationally scalable for large-scale datasets. Table 2.2 summarizes the main research areas, the dominant approaches in the literature, their limitations, and the specific gaps addressed by this thesis.

Table 2.2: Synthesis of State-of-the-Art Limitations and Identified Research Gaps

RQ	**Analytical Objective**	**Methods and Artifacts**	**Main Results and Evidence**
RQ1	Data acquisition and standardization	Data preprocessing, semantic normalization, and quality control of event logs	Heterogeneous clinical data successfully transformed into structured and consistent event logs, enabling reliable downstream analysis
RQ2	Process reconstruction and pathway analysis	Process discovery algorithms, activity reconstruction, and enhanced event logs	Care pathways reconstructed despite high variability, allowing dominant patterns and deviations to be identified
RQ3	Temporal and behavioural analysis	Extraction of temporal metrics, activity frequencies, and transition analysis	Temporal and behavioural features reveal significant differences in patient trajectories and process complexity
RQ4	Predictive and adaptive modelling	Supervised machine learning models enriched with process-derived and temporal features	Improved predictive performance compared to baseline models, with robustness to class imbalance and temporal variation
RQ5	Governance and continuous maintenance	Drift detection, model monitoring, and human-in-the-loop validation	Sustained model reliability achieved through adaptive maintenance and transparent governance mechanisms

The synthesis presented in Table 2.2 highlights that existing approaches tend to address individual aspects of healthcare analytics in isolation, often neglecting the interdependencies between data quality, process dynamics, predictive modelling, and long-term governance. In particular, the lack of integrated solutions capable of adapting to evolving clinical processes emerges as a recurring limitation across the literature.

In conclusion, this chapter has outlined the limitations of the state of the art, identifying

the inverse relationship between structural complexity and model stability as the primary hurdle for current solutions. The original contribution of this thesis lies in overcoming these limitations through: (i) automated semantic reconciliation (bridging the gap of data fragmentation), (ii) structural complexity reduction techniques (addressing the issue of structural over-complexity in clinical models), and (iii) a continuous monitoring layer (filling the gap in clinical governance).Based on this critical assessment, the following chapter introduces the definition of the proposed framework.

CHAPTER

3 Framework Definition and Planning

3.1 Healthcare Process Mining Maintenance Framework

The growing availability of healthcare data and the increasing adoption of Process Mining techniques have created new opportunities for analyzing complex clinical processes. However, most existing approaches focus on static analyses performed on historical data, often neglecting the challenges related to model sustainability and long-term maintenance in dynamic healthcare environments.

To address this limitation, this research proposes the Healthcare Process Mining Maintenance Framework, a structured methodology designed to support the full lifecycle of healthcare process mining and predictive analytics. The framework integrates data acquisition, data quality management, process discovery, predictive modeling, and governance mechanisms within a unified analytical ecosystem.

The following sections introduce the conceptual foundations of the framework, define its operational boundaries and constraints, and describe the layered architecture that supports continuous model maintenance.

3.1.1 Conceptual Overview of Framework

The proposed framework is designed to provide a unified structure capable of integrating the critical components required for healthcare process mining and predictive modeling. Its

primary objective is to transform heterogeneous clinical data into reliable analytical models while ensuring their continuous validation and adaptation over time.

Within this perspective, maintenance is not treated as an auxiliary activity but rather as a core methodological principle embedded within the analytical architecture. The framework therefore supports both the generation and the continuous evolution of analytical models in response to changes in clinical processes and data distributions.

Formally, the proposed framework can be defined as a Continuous Adaptive System:

$$\mathscr{S} = \langle \mathscr{D}, \mathscr{M}, \mathscr{G}, \phi \rangle$$

where:

- $\mathscr{D}$ represents the dynamic data space, including clinical events, electronic health records, and event logs;
- $\mathscr{M}$ denotes the set of process models and predictive models derived from the data;
- $\mathscr{G}$ represents the governance layer, which includes institutional policies, clinical guidelines, and regulatory constraints;
- ϕ denotes the maintenance function responsible for updating the system in response to data evolution.

More specifically, the maintenance function can be interpreted as a dynamic transformation that maps the current state of data and models into an updated system configuration:

$$\phi : (\mathscr{D}_t, \mathscr{M}_t) \rightarrow (\mathscr{D}_{t+1}, \mathscr{M}_{t+1})$$

This formulation reflects the adaptive nature of the framework, where both data representations and analytical models may evolve over time. Unlike traditional linear analytical pipelines (Data $\rightarrow$ Model $\rightarrow$ Deployment), the proposed framework incorporates a recursive feedback mechanism in which monitoring and governance activities continuously trigger updates to data processing and modeling components.

3.1.2 System Boundaries and Operational Constraints

To strengthen the conceptual foundation of the framework, it is necessary to explicitly define its systemic boundaries and operational constraints. This formalization ensures that the framework operates as a bounded analytical system rather than as a simple sequence of analytical steps.

System Boundaries The framework distinguishes between the *External Environment* and the *Internal Analytical System*. The external environment includes clinical information systems such as Electronic Health Records (EHR), Picture Archiving and Communication Systems (PACS), laboratory platforms, and administrative databases. These systems generate the raw data that serve as the input for analytical processes.

The internal analytical system corresponds to the Healthcare Process Mining Maintenance Framework itself. Its operational boundary begins with the Data Acquisition and Standardization layer, which transforms heterogeneous clinical data into structured event logs, and extends to the Governance and Continuous Maintenance layer, which supervises model performance and ensures regulatory and clinical compliance.

Operational Constraints The maintenance cycle implemented by the framework operates under a set of fundamental constraints designed to ensure analytical reliability and clinical interpretability.

- **Data Fidelity Constraint:** Analytical models can only be updated after the completion of a data quality assessment and repair phase. This constraint prevents the propagation of errors or inconsistencies present in raw clinical data.

- **Structural Parsimony:** Process discovery techniques must maintain a balance between accuracy and interpretability. When discovered models exceed a predefined complexity threshold, model simplification or abstraction procedures are applied to ensure their interpretability for clinical stakeholders.

- **Governance Override:** The framework adopts a human-in-the-loop approach. Significant model updates or structural modifications must be validated through governance procedures before being operationally adopted.

These constraints ensure that the analytical lifecycle remains aligned with both technical performance requirements and clinical governance principles.

3.1.3 Conceptual Architecture

The conceptual architecture of the Healthcare Process Mining Maintenance Framework is illustrated in Figure 3.1. The architecture adopts a layered structure composed of five interconnected analytical levels that collectively support the lifecycle of healthcare process mining and predictive modeling.

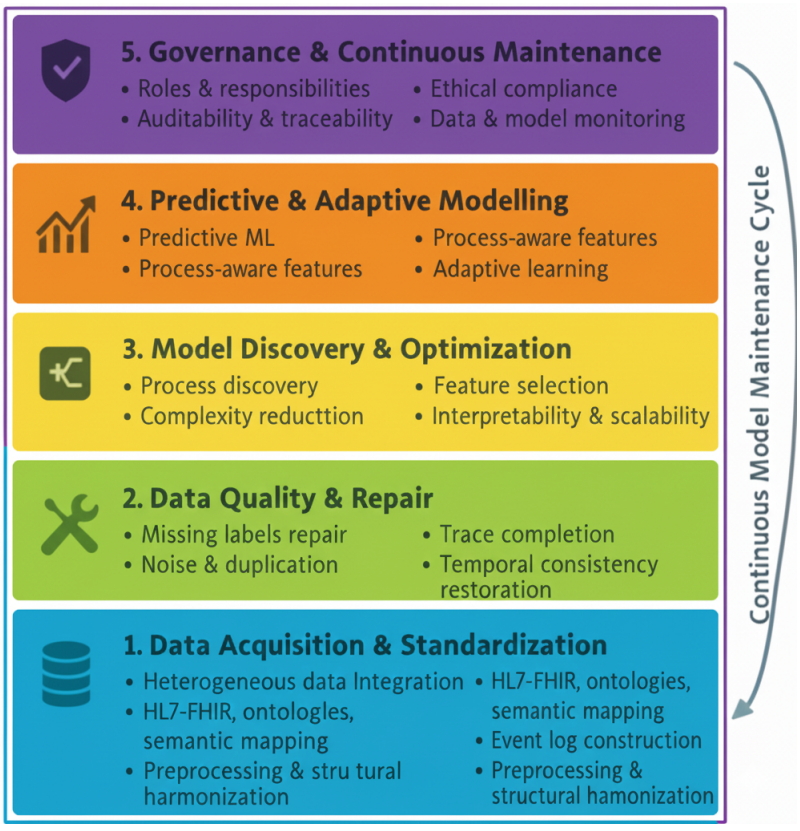

Figure 3.1: Healthcare Process Mining Maintenance Framework Conceptual Architecture.

The framework is organized according to a hierarchical structure in which each layer

performs a specific analytical function while simultaneously interacting with the others through iterative feedback mechanisms. This organization reflects the inherently dynamic nature of healthcare processes, where analytical models must continuously adapt to evolving clinical practices and data patterns.

The five layers composing the framework are:

- **Layer 1 – Data Acquisition & Standardization**, responsible for integrating heterogeneous clinical data sources and transforming them into standardized event logs suitable for process mining analysis. This layer acts as the semantic gateway and normalization interface of the framework. This layer acts as the semantic gateway and normalization interface of the framework. Its primary objective is to minimize information entropy during data extraction while preserving the clinical context of recorded events. This directly addresses **RQ1** (*How can heterogeneous clinical data be effectively acquired and standardized?*) by providing a formal mapping between raw sources and a unified, interoperable event log structure.

- **Layer 2 – Data Quality & Repair**, which focuses on detecting and correcting inconsistencies, missing information, and behavioral noise within event logs. This layer introduces a proactive data-repair loop within the framework. By restoring the fidelity of event logs through probabilistic alignment and pattern matching, it ensures the reliability of the evidence base. This stage specifically responds to **RQ2** (*How is data quality ensured in clinical event logs?*) by shifting data quality management from a one-time step to a continuous maintenance requirement.

- **Layer 3 – Model Discovery & Optimization**, dedicated to the extraction of interpretable process models capable of representing complex healthcare workflows. Acting as the interpretability engine, this layer focuses on mitigating structural over-complexity (the ”spaghetti” effect) in healthcare workflows. By transforming high-density process graphs into human-readable pathways, it provides the formal representation required to answer **RQ3** (*How can complex healthcare processes be represented in an interpretable and scalable way?*).

- **Layer 4 – Predictive & Adaptive Modeling**, which integrates process mining features with machine learning techniques to generate predictive insights and detect process drift.This component represents the proactive intelligence core of the framework. Its objective is to maintain predictive performance through automated monitoring of performance degradation. It directly addresses **RQ4** (*How can predictive models be trained and maintained?*) by ensuring that predictive intelligence remains sustainable despite the evolution of clinical practices.

- **Layer 5 – Governance & Continuous Maintenance**, responsible for monitoring model performance, ensuring regulatory compliance, and coordinating the long-term sustainability of the analytical ecosystem.This final layer defines the feedback policy and systemic boundaries of the framework. By formalizing clinical auditability and institutional alignment through human-in-the-loop validation, it establishes the trustworthy environment required to answer **RQ5** (*How can process mining systems remain trustworthy, auditable, and aligned with governance principles?*).

Together, these layers form an integrated analytical pipeline in which data preparation, model generation, predictive analysis, and governance activities operate within a continuous maintenance cycle. This layered architecture ensures both methodological coherence and operational sustainability in real-world healthcare environments.

3.2 Data Acquisition & Standardization

The success of any Process Mining and Machine Learning activity within the healthcare domain is critically dependent on the quality, consistency, and completeness of available clinical data. Clinical environments, however, produce information through a large and diverse set of information systems: electronic medical records, triage systems, laboratory platforms, radiology systems, administrative archives, and many other sources distributed across the hospital organization. The diversity of data structures, recording protocols, and temporal granularities makes any form of integrated analysis complex. This challenge is compounded by terminological differences, semantic duplications, divergent encodings used by departments,

and the lack of common standards for recording clinical activities.Any errors, omissions, or gaps introduced at this foundational stage will inevitably propagate through subsequent analyses, impacting discovery, predictive modeling, and the validity of conclusions.

This critical gap in data quality and consistency naturally leads to the research question :

RQ1: *How can heterogeneous clinical data be effectively acquired and standardized for process mining?*

For this reason, the framework recognizes the need for preliminary intervention in the data acquisition and standardization phase. The process adopted aims to answer the key question of how to obtain a robust and interoperable event log from structurally heterogeneous data. The first step involves systematically identifying relevant information sources, distinguishing between data required for process reconstruction, contextual data, and metadata useful for subsequent feature creation. Once identified, the sources are subjected to a conceptual mapping process toward standard models such as HL7-FHIR and clinical ontologies, ensuring semantic consistency and reducing ambiguity. This is followed by structural harmonization, which includes format normalization, resolution of temporal discrepancies, encoding unification, and event synchronization. Only then can we proceed to construct the event log, which constitutes the central representation for subsequent analyses. The event, in its final form, must be reliable, consistent, and, above all, reusable for modeling, prediction, and evaluation activities.This layer represents the foundational stage of the framework: any errors, omissions, or gaps introduced at this stage will inevitably propagate through subsequent analyses, impacting discovery, predictive modeling, and the validity of conclusions.

This layer manages the extraction and harmonization of heterogeneous clinical data sources, including EHRs, laboratory systems, and administrative databases. It ensures semantic interoperability through mapping to standardized formats such as HL7-FHIR and domain ontologies, thereby creating the foundation for reproducible and comparable process analyses.

The framework's design principles were shaped by the integration challenges observed across the empirical studies, particularly when working with heterogeneous datasets such as MIMIC-IV and the Sepsis cohort, each characterized by different structures, levels of granularity, clinical workflows, and documentation practices. These differences required the adoption

of different preprocessing strategies to ensure that the data could be analyzed in a coherent and comparable manner. The necessity to reconcile these heterogeneous preprocessing techniques, while preserving the semantic integrity of each dataset, highlighted the importance of a unified and modular framework capable of accommodating dataset-specific requirements without losing methodological consistency. As a result, the framework was designed to systematize the integration of heterogeneous clinical data and support a repeatable, maintainable pipeline across varied real-world settings.

3.3 Data Quality & Repair

While the previous phase (RQ1) ensures that data is acquired, structured, and represented correctly, a critical issue remains: the actual quality of clinical logs. Empirical experience and the literature demonstrate that, in healthcare settings, process recording issues are not sporadic but systematic. Unlike many administrative or industrial processes, clinical processes are documented under operating conditions characterized by high care pressure, heterogeneous information tools, and priority given to clinical action over digital recording. This results in logs affected by semantic incompleteness, missing or incorrect labels, inconsistent timestamps, incorrectly reported concurrent activities, duplications, and a significant level of behavioral noise.

These critical issues are not simply formal flaws but have direct implications for the validity of analyses. Studies, particularly those dedicated to recovering missing activities using machine learning models constrained by process rules, show how semantic incompleteness compromises all subsequent phases of the pipeline. Patterns discovered from unrepaired logs are inaccurate or distorted: seemingly anomalous patterns often arise from data gaps rather than actual clinical deviations; discrepancies from protocols tend to be amplified or misinterpreted; and predictive models struggle to recognize clinically significant signals because they are based on traces that do not reflect the actual behavior of the process.

This critical need for reliable representation leads directly to the second research question:

RQ2: *How is data quality ensured in clinical event logs?*

To address this need, the framework proposes a dedicated and structured phase for data quality assessment and repair. This phase aims to ensure that the event log is a faithful, consistent, and semantically complete representation of the underlying clinical process. The work performed in this phase is not limited to correcting superficial errors but also deeply impacts the semantics of the process,specifically, the study involves:

- Systematic identification of gaps through completeness, consistency, and compliance analyses with reference models.
- Reconstruct missing activities using predictive models trained on complete traces, combining behavioral information and process rules.
- Restore temporal consistency through alignment, smoothing, and timestamp correction techniques, especially in cases of simultaneous or reversed recordings.
- Manage duplications and spurious activities with targeted noise filtering and reconciliation of inconsistent sequences.
- Preserve process semantics, avoiding corrections that alter clinical logic or introduce artificial pathways that are inconsistent with guidelines or observed experience.

The ultimate goal is not simply to produce a ”clean” log but to provide a log that realistically, consistently, and semantically informedly reflects the behavior of the clinical process. Only with repaired and reliable logs is it possible to rigorously initiate the subsequent phases of process discovery, behavioral analysis, and predictive modeling.

3.4 Model Discovery & Optimization

Once the clinical event log has been acquired and reliably repaired (via RQ1 and RQ2), the focus shifts to extracting actionable insights. This phase represents the crux of Process Mining in the clinical setting: the creation of process models that are not only faithful to the observed reality but, crucially, are also interpretable and directly suited to the practical needs of the healthcare context. This requirement arises from the complexity of care pathways,

characterized by nonlinear flows, significant variability between patients and within the same patient's journey, as well as the presence of numerous repetitive activities or those influenced by emergency factors.

Experience suggests that traditional single-technique approaches are insufficient to manage this complexity while maintaining both fidelity and interpretability. The guiding question for this phase is therefore how to achieve models that are both representative and manageable, striking an optimal balance between faithfulness to the real process and clarity for clinical interpretation.

RQ3 : *How can complex healthcare processes be represented in an interpretable and scalable way?*

To effectively address this complexity, experience suggests that a single technique is not sufficient; a targeted combination of interventions is required. The first fundamental intervention involves reducing the log size, achieved through stratified sampling. Rather than randomly reducing the data volume, this strategy allows for the selection of a subset of cases (or traces) that, despite being smaller in number, retain the full behavioral variety observed in the original log. This allows for a smaller, more manageable log, while maintaining the representation of crucial, albeit less frequent, pathways.

The second study relies on temporal aggregation, which directly addresses the problem of noise and excessive granularity. Activities that repeat closely and rapidly, typical of the clinical environment (such as serial measurements or monitoring), are grouped into a single macro-event or synthesized. This process has the effect of cleaning the model, drastically reducing the number of non-informative nodes and edges, which significantly improves the readability and interpretability of the resulting graph.

Finally, temporal feature engineering establishes the crucial link between process discovery and predictive analytics. The sequential and temporal information extracted from the cleaned models (such as the time between a diagnosis and the start of treatment, or the frequency of certain key activities) is transformed into high-level features. These variables become robust and meaningful inputs for machine learning algorithms. The quality of the discovered models, being compact and stable, directly impacts the robustness of these features, ensuring

better generalization and supporting more reliable forms of advanced analysis, essential for predicting clinical outcomes or process deviations.

This phase of the framework produces models that represent real-world processes more clearly and with greater interpretability, but also form the basis for generating the high-level features needed for predictive modeling. The quality of the discovered models directly impacts the ability to predict clinical outcomes or process deviations. A more compact, stable, and interpretable model ensures better generalization and supports more reliable forms of advanced analysis.

3.5 Predictive & Adaptive Modelling

The fourth level of the framework is the analytical component most oriented toward prediction and adaptivity. In healthcare, the development of predictive models faces unique challenges related to the highly dynamic nature of clinical processes: patient trajectories evolve over time, care protocols are updated, and department operating conditions vary based on organizational and emergency factors. In this context, obtaining models with high predictive performance is only part of the challenge. The critical issue lies in preventing the rapid model decay that typically occurs in such dynamic environments. Therefore, the core problem is designing systems capable of continuously and autonomously adapting to changes in data and process behaviors over time.

RQ4: *How can predictive models be trained and maintained to anticipate clinical outcomes?*

The empirical evidence provided by the [**emergency˙care˙study˙MIMICEL**] confirms the value of process-derived features, particularly temporal characteristics, in predicting clinical outcomes. The approach adopted involves building a predictive pipeline that integrates heterogeneous information: activity sequences, temporal indicators (durations, inter-event times, time-to-event), metrics extracted from process models (frequencies, transitions, compliance), and patient-specific clinical variables. This integration enables the capture both the information content of traditional clinical data and the dynamic and behavioural components inherent in care pathways. Once built, the models are constantly monitored using

drift detection techniques, which identify significant variations in data distribution or process behavioural dynamics.

When necessary, the models are updated or retrained, ensuring stable performance over time. This component of the framework is essential for real-world applications, as it prevents the rapid model decay that typically occurs in dynamic contexts such as healthcare. The predictive modeling layer therefore represents not only an advanced analysis phase, but a true adaptive capability of the framework, strengthening its robustness over time.

From this perspective, the fourth level consolidates the overall robustness of the framework, translating the principles of continuous maintenance into a concrete and integrated analytical practice.

3.6 Governance & Continuous Maintenance

The final level of the framework focuses on the organizational, managerial, and ethical dimensions essential to ensuring the sustainability of the entire analytics ecosystem. Even the most accurate predictive or descriptive models quickly lose value if they are not embedded in an appropriate governance framework equipped with oversight, maintenance, and control mechanisms. The central question guiding this level is therefore how to ensure transparency, reliability, and continuity in the use of data and models over time, while maintaining adherence to clinical protocols, institutional policies, and current regulations.

In this perspective, the governance layer is designed to bridge the gap between technical performance and regulatory compliance, specifically aligning with the requirements for high-risk AI systems established by the EU AI Act. By implementing systematic logging and performance tracking, the framework ensures the traceability (Art. 12) and human oversight (Art. 14) necessary for clinical decision-support tools.

Unlike the other levels, this phase does not stem from a single empirical contribution but represents the synthesis of socio-technical requirements that have emerged across all previous phases. The framework explicitly assumes that maintenance is not a mere technical activity but rather an ongoing organizational process, requiring the definition of roles, responsibilities,

and institutional procedures. These include model performance oversight, auditability and traceability of changes, definition of version and update policies, ongoing data monitoring, periodic model evaluation, drift management, and documentation of the impact of analyses on clinical and decision-making processes.

This essential requirement for long-term sustainability and ethical integration leads us to the fifth research question:

RQ5: *How can process mining systems remain trustworthy, auditable, and aligned with clinical governance principles?*

Furthermore, the framework facilitates policy-level translation by converting technical drift indicators into actionable management insights. For example, a sustained decrease in model reliability caused by changing patient demographics can reported for a potential shift in care demand, enabling administrative adjustments in resource allocation or clinical protocol updates.

The governance level therefore ensures that the framework is not a fragmented set of techniques but rather a coherent, structured, and reproducible system integrated with the objectives and practices of the healthcare context. This prevents the risk of models becoming obsolete, being used outside their scope of validity, or producing unexpected effects on care pathways. At the same time, it ensures the long-term sustainability of the analytical infrastructure by promoting responsible, transparent, and robust use of data-driven systems.

3.7 Continuous Model Maintenance Cycle

The architecture highlights a deep functional interdependence between the layers, signaled by the bidirectional flow: Data Quality (L2) is the fundamental prerequisite for effective Model Discovery (L3). The structural knowledge of the Process Model (L3) informs and enhances the training of the Predictive Model (L4). The accuracy of the predictions (L4) and their clinical implications inform governance decisions (L5). Governance (L5), through audit and monitoring, determines the new quality and standardization requirements (L1/L2) and the need for model retraining (L4), closing the loop. This recursive logic, represented by the

Continuous Model Maintenance Cycle show in Figure 3.2, establishes the framework not as a linear process but as a dynamic and self-regulating ecosystem that constantly evolves in tune with data and changing real-world clinical processes. The conception of the Healthcare Process Mining Maintenance Framework emerged from the recognition of a fundamental gap between the increasing methodological maturity of process mining techniques and their long-term sustainability within real-world healthcare environments. Over the last decade, process mining has progressively demonstrated its potential to extract actionable insights from Electronic Health Records (EHRs), from event logs and other clinical information systems. However, despite this growing interest, most implementations have remained confined to experimental or pilot contexts.

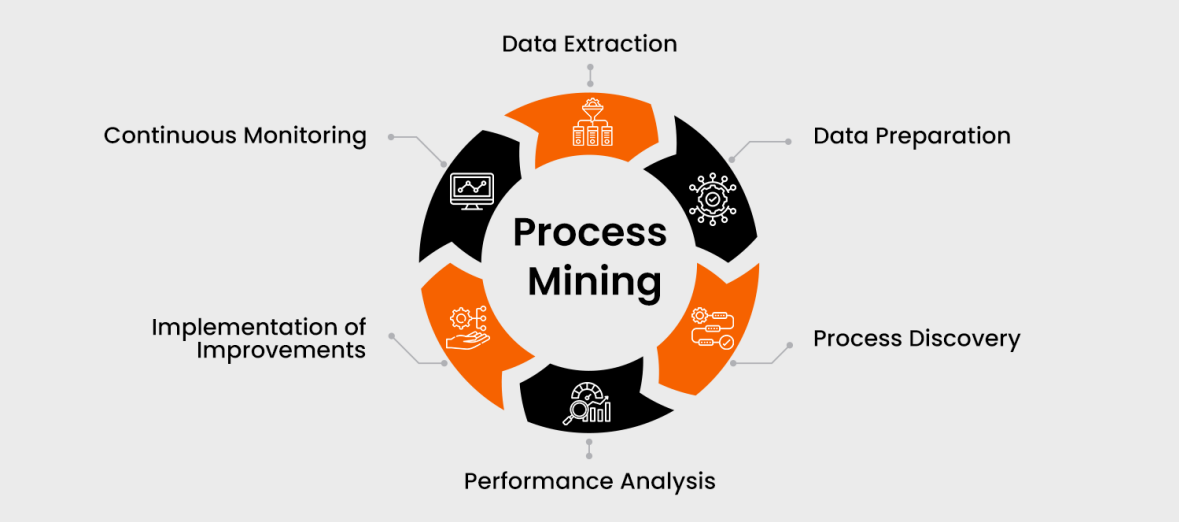

Figure 3.2: Continuous Model Maintenance Cycle.

The main obstacle has not been the lack of analytical sophistication, but rather the difficulty of ensuring that models remain valid, interpretable, and aligned with evolving data and clinical practices. The maintenance cycle transforms the framework from a static analytical pipeline into a continuous adaptive system, where monitoring and governance mechanisms trigger updates across all analytical layers.

CHAPTER

4 Framework Structure and Implementation

4.1 Overview of the Experimental Pipeline

The experimental methodology is organized through a modular pipeline designed to bridge the gap between theoretical framework requirements and clinical data complexity. Unlike standard process mining implementations, this pipeline is structured as a closed-loop system where each phase is specifically engineered to address a distinct Research Question (RQ) through targeted data transformations.

The pipeline is organized into five main functional blocks:

Data Engineering (RQ1) This block handles the physical ingestion and standardization of raw clinical data. Unlike standard ETL processes, it implements a custom semantic mapping to ensure interoperability between heterogeneous sources.

Log Repair Engine (RQ2) This module implements a probabilistic reconstruction algorithm. Its purpose is to transform fragmented logs into semantically complete traces, mitigating the ”behavioral noise” typical of clinical documentation.

Structural Optimization (RQ3) To manage process complexity, this stage applies stratified sampling and temporal aggregation. The technical goal is to produce process models that optimize the trade-off between structural fidelity and human interpretability.

Adaptive Predictive Pipeline (RQ4) This final block manages feature engineering and the training of adaptive machine learning models, integrating a monitoring layer to detect performance degradation over time .

Monitoring & Maintenance Logic (RQ5) This final module implements the governance layer. It provides the technical triggers for model retraining and audit trails, ensuring that the system remains trustworthy and aligned with clinical protocols through continuous performance monitoring.

While each component of the pipeline is based on established techniques from process mining and machine learning, the originality of the proposed framework lies in the integration and adaptation of these components for the clinical process analysis domain. This framework enables the transformation of fragmented clinical logs into predictive-ready datasets while preserving process semantics, which represents the main methodological contribution of this work.

4.2 Data Sources and Preprocessing Procedures

4.2.1 Public Event Logs

Several public process logs were used to evaluate the predictive models and analyze temporal process characteristics.

The datasets considered come from the Business Process Intelligence Challenge (BPIC) 2012 [17] and BPIC 2017 [18]:

- BPIC 2012 captures the loan application process of a Dutch financial institution from October 2011 to March 2012, including 13,087 cases and 262,200 events with 36 unique activity labels. To facilitate the analysis, three versions were extracted based on the intersection of resource and event status information.

- BPIC 2017, from the same institute but with a different system, covers events from 2016

to February 2017, comprising 31,509 cases, 1,202,267 events, and 26 unique activity labels.

These datasets were selected both for their diverse characteristics (activity count, trace complexity, and temporal context) and for their widespread use as benchmarks in the literature. This allows for direct comparison with previous studies, providing a solid basis for validating the methodology.

The datasets offer significant sequential richness, with process paths that diverge substantially across traces, enabling the evaluation of predictive techniques in different contexts and non-uniform behavior. Differences in temporal granularity and activity frequency further impact feature extraction and model robustness.

Finally, the use of publicly available registries ensures comparability with previous work, establishing a reliable benchmark for evaluating the performance of preprocessing procedures and predictive models, while aligning the study with recognized best practices in process analysis.

4.2.2 Clinic Event Logs

Clinical event log analysis is one of the most significant methodological challenges in process mining. Data from healthcare systems reflect the non-linear, stochastic, and dynamic nature of care pathways. Unlike standardized operational processes common in industrial sectors, hospital workflows exhibit high variability. Patient heterogeneity and unpredictable therapeutic responses generate complex and often fragmented execution traces.

The need to repair logs arises from fragmented hospital information systems and critical operating conditions of medical staff. These factors often prevent timely and synchronous activity recording. Information gaps show up as missing key events or disorganized timestamps. Reconstruction techniques and alignment algorithms, such as conformance checking, deduce missing events by comparing incomplete traces to clinical protocols. This process ensures logical consistency and transforms raw data into a robust analytical model.

The effectiveness of these techniques has been tested on public reference datasets, such as:

- Hospital Billing: a log extracted from regional ERP systems relating to medical billing processes, characterized by a high volume (100,000 records).

- Sepsis Registry: a registry focused on the clinical management of patients with sepsis, where the accuracy of the temporal sequence is essential for analyzing the therapeutic pathway.

Aggregation and sampling strategies address the critical issues posed by data dimensionality and granularity. In a clinical context, a single hospitalization episode can generate thousands of micro-events (e.g., continuous monitoring or drug administration). When analyzed atomically, these data produce so-called spaghetti models, excessively dense graphs devoid of information value.

Aggregation allows elementary events to be grouped into clinically significant macro-phases, while sampling ensures the computational sustainability of the analysis. A prime example is the use of the MIMIC-IV-ED dataset, which includes over 7.5 million events related to emergency department flows. In this scenario, the proposed approach allows us to isolate critical behaviors and deviations from standard pathways without being obscured by the background noise of routine data.

Ultimately, the normalization process transforms a heterogeneous set of administrative and clinical records into a structured and rigorous model. This refinement is the fundamental prerequisite for extracting knowledge aimed at optimizing resources and improving patient safety, ensuring that each digital trace faithfully corresponds to the actual care experience delivered.

4.2.3 Emergency Department Event Logs

originally structured as time sequences of events, into a format suitable for advanced computational analysis. These logs reflect the entire patient journey, including critical phases such as admission, triage, vital signs monitoring, diagnostic and therapeutic procedures, and final outcome. The core of the trial uses the MIMICEL dataset, derived from the MIMIC-IV-ED repository, which represents the end-to-end process of patient flows in the emergency

setting. In its raw form, the dataset comprises over seven million events distributed across approximately four hundred and twenty-five thousand cases, providing an exceptionally broad empirical basis for optimizing operational efficiency.

To make this data compatible with machine learning architectures, the pipeline includes a complex restructuring operation that converts the sequence of events into a structured tabular representation. At the end of this process, the dataset consists of 389,681 instances and 18 predictor variables. In this new configuration, each row identifies a single patient journey, integrating static variables, such as age and comorbidities, with dynamic information aggregated from time logs, including waiting times between activities and statistical averages of vital signs recorded during the patient's stay.

The target variable selected for the classification tasks is "Disposition," an indicator that encodes the patient's exit status through eight distinct categories, such as discharge home or voluntary departure. This variable is crucial for modeling patient flows and predicting the intensity of care required after the emergency phase. The transition from sequential logs to structured datasets, therefore, represents the methodological cornerstone of the work, as it allows for the preservation of the informative richness of clinical data while ensuring compatibility with supervised learning models. Finally, this approach allows for rigorous management of missing data and variable normalization, providing robust predictive tools to support clinical and organizational decisions.

4.2.4 Preprocessing

The preprocessing stage follows a structured workflow composed of three main phases:

1. **Input**: raw event logs extracted from process information systems and clinical repositories.

2. **Processing**: data cleaning, temporal ordering, feature extraction, and semantic normalization.

3. **Output**: structured datasets suitable for predictive modeling and process analysis.

This structure ensures reproducibility and clarity regarding the transformation of raw logs into analytical datasets. In the following sections, the preprocessing operations corresponding to the processing phase are described in detail for both process event logs and emergency department clinical logs.

All logs used underwent a preprocessing phase consisting of several sequential operations, with the aim of purifying, normalizing, and transforming the raw data into a coherent format suitable for predictive modeling. Although the methodological framework is common, the processing was differentiated to address the specificities of process event logs and emergency department clinical logs.

For process event logs, the fundamental unit of analysis is the trace, defined as an ordered sequence of events $T = \langle e_1, e_2, \ldots, e_n \rangle$ pertaining to the same case. Initially, the events within each trace were sorted chronologically; subsequently, the entire corpus of traces was sorted based on the timestamp of the first event of each instance. To predict the next activity, a prefix extraction technique was adopted using a sliding window of size $w = 3$. This value represents a compromise between contextual richness and model complexity. Preliminary exploratory analysis showed that shorter windows ($w = 1$ or $w = 2$) were insufficient to capture meaningful behavioral patterns, while larger windows significantly increased dimensionality without improving predictive performance. Therefore, $w = 3$ was selected as a balanced configuration widely adopted in next-activity prediction tasks. This approach generates sequences of the form $\langle e_{k-w}, \ldots, e_k \rangle$ to predict the event e_{k+1}. In cases where the position k is smaller than the window size, a zero-padding strategy was applied to ensure dimensional uniformity of the input. Concurrently, a systematic temporal feature extraction was performed along three dimensions: the latency between consecutive events, the time of day, and the day of the week. The temporal difference was calculated as $\Delta t = t_k - t_{k-1}$, setting the value to zero for the initial or padding events. The hour $[0, 23]$ and the day of the week $[0, 6]$ were numerically encoded and concatenated to the activity prefixes. Finally, the activity labels were label-encoded, mapping each activity to a discrete interval $[0, N-1]$, making the dataset compatible with machine learning algorithms. This encoding strategy was selected because it preserves the categorical nature of activities while ensuring compatibility with standard

machine learning algorithms that require numerical input representations.

Preprocessing the Emergency Department clinical logs required specific interventions to manage the heterogeneity of unstructured data. Initially, a semantic normalization process was performed on the textual variables, with particular attention to the chief complaint. Three information dimensions were extracted from the latter: symptoms, anatomical region involved, and primary etiology. This mapping process allowed for the standardization of synonyms and lexical variants (for example, associating ”hypertension” and ”high blood pressure” with a single concept), drastically reducing the dimensionality of the dataset. Data quality was further ensured by removing outliers and clinically implausible values for vital signs. Thermal measurements were standardized into a single unit of measurement, while qualitative descriptions of pain were converted into a continuous numerical scale from 0 to 10 using semantic mapping. Finally, to preserve model robustness, a rigorous analysis of missing values was performed, removing observations with gaps in critical variables. This data preparation protocol ensures that subsequent analyses are not influenced by statistical noise or semantic inconsistencies, laying the foundation for accurate and clinically relevant predictive capability.

4.3 Log Quality Evaluation and Repair

Clinical event logs are frequently incomplete due to delayed or missing data registration. To address this issue, a dedicated Log Repair Engine was implemented.

Input Fragmented event traces extracted from clinical information systems.

Processing Snippet-based contextual analysis using bidirectional sliding windows to detect inconsistencies and infer missing activities.

Output Semantically reconstructed traces suitable for downstream process mining and predictive modeling.

The implementation of the Log Repair Engine introduces a customized snippet-based reconstruction logic. The originality of this approach lies in the use of a bidirectional sliding

window that incorporates both preceding and succeeding clinical events to infer missing nodes. Unlike standard imputation methods, the engine handles trace boundaries through a null-event padding strategy, preventing the algorithm from generating artificial process paths at the beginning or end of a patient's stay.

4.3.1 Detection of Log Incompleteness

Analyzing clinical processes requires a preliminary transformation of the raw data into structures that allow us to observe the context of each event. A key technique for achieving this is Process Snippet Extraction, a feature engineering methodology that enriches each activity in the log with information about its past and immediate future. Specifically, for each recorded activity, additional variables called PREACTIVITY and NEXTACTIVITY are generated. These functions act as a sliding window of variable width w (with values from 1 to 4), capturing the sequence of events before and after position k in a trace T. Formally, given an ordered trace:

$$t = \langle \#activity(e_1), \#activity(e_2), \ldots, \#activity(e_i), \ldots, \#activity(e_n) \rangle$$

the value of preActivity for $w = 1$ corresponds to $\#activity(e_{i-1})$. If multiple previous activities are evaluated, the process is reversed: $preActivity_2 = \#activity(e_{i-2})$, $preActivity_3 = \#activity(e_{i-3})$, and so on. The same method is applied to define the nextActivity, looking at subsequent events: $nextActivity_1 = \#activity(e_{i+1})$ up to $nextActivity_4 = \#activity(e_{i+4})$. This process fragment(snippet) structure is the essential tool for the subsequent incompleteness identification phase. Since clinical logs are often fragmented, a targeted analysis of these fragments allows us to highlight three types of critical issues:

- Missing activities: identified when the extracted patterns show unjustified gaps between events that, according to the medical protocol, should be consecutive.
- Chronological anomalies: detected by observing the temporal sequences in the snippets, as in the case of reports recorded before the exam itself.
- Deviations: highlighted by comparing the extracted snippets with the expected behavior

from standard clinical pathways.

A crucial technical aspect concerns the management of the trace boundaries. For a sequence defined as $t = \langle \#activityStart, \#activity(e_1), \ldots, \#activity(e_n), \#activityEnd \rangle$, it is necessary to handle cases where the window w extends beyond the beginning or end of the pathway. Two options were considered: assigning a common label ”NoActivity” for both cases, or distinguishing between the beginning and the end using the labels ”NoPrevious” and ”NoNext.” This uniformity allows the algorithm to correctly distinguish between an absence due to a registration error and one physiologically due to the beginning or end of the patient's pathway. Ultimately, the ability to detect clinical incompleteness directly depends on the chosen window depth: a wider window ($w = 4$) offers a richer view, making it easier to recognize complex irregularities that would otherwise be missed by observing individual events in isolation.

4.3.2 Reconstructing Missing Activities

Once the detection phase has isolated an anomaly or gap in the sequence, the next step is reconstruction. Instead of simply eliminating incomplete traces (losing valuable information), the system uses the structure defined in Snippet Extraction to fill in the gaps. The model acts as a predictive system: by analyzing the *preActivity* and *nextActivity* features within a windows w, it estimates which activity is most likely to close the broken pattern.This approach does not invent data, but infers the omitted event based on the behavior observed in thousands of other similar traces, ensuring that the final trace is consistent with reality.

4.3.3 Validation

Inserting reconstructed activities carries the risk of introducing noise or statistical bias. For this reason, the validation phase is essential to ensure that the resulting log faithfully reflects clinical practice. This verification is not limited to the accuracy of the individual inserted activity, but analyzes the entire dataset under three aspects:

- Internal consistency: It verifies that the new sequences do not violate medical or logical constraints (for example, a discharge can never be inserted before an operation if the

reconstruction occurs between these two points).

- Frequency analysis: The statistics of the original log are compared with the reconstructed one. If a specific activity suddenly appears with a disproportionate frequency compared to the raw log, the reconstruction process may have introduced a bias.
- Transition stability: It is observed whether the probabilities of switching from one activity to another have remained stable.

The goal is for the repaired log to improve process flow without altering the actual behavior initially observed. This rigorous review phase transforms the log from a fragmented collection of data to a reliable knowledge base, ready for more complex discovery algorithms or compliance checks with healthcare guidelines.

4.4 Structural Complexity Reduction

To address the 'spaghetti model' problem, the implementation employs a Stratified Feature-Based Sampling. The distinctive element here is that the sampling is not random; it is guided by the preservation of 'variant entropy'. This ensures that even rare clinical pathways (the 'long tail' of the distribution) are retained in the final model, allowing the optimization process to reduce visual noise without losing the clinical edge cases that are often the most critical for maintenance and audit.

4.4.1 Stratified Sampling

The size and complexity of clinical logs often necessitate reducing the amount of data to be analyzed to make the subsequent discovery and analysis phases computationally manageable. To address this need, a stratified sampling strategy is employed that goes beyond random selection, but instead divides the traces based on two key dimensions: trace length and variant frequency. This approach allows for maintaining the original behavioral diversity while significantly reducing the data volume. Regarding Layering by Length, given a trace σ_i with length $l(\sigma_i) = m$, the original log L is split into three disjoint sets based on thresholds (θ_1, θ_2)

defined by the quantiles of the distribution:L_{short}: traces with $l(\sigma) \leq \theta_1$.L_{medium}: traces with $\theta_1 < l(\sigma) \leq \theta_2$.$L_{long}$: traces with $l(\sigma) > \theta_2$. This split is crucial in the clinical setting because very short traces often represent simplified or interrupted processes, while long traces describe complex care pathways or chronic patients. Once the three ”bins” (short, medium, long) have been defined, it is necessary to decide how many traces to take from each to reach the total target N. Two sample allocation methods were considered: Equal allocation: the sample is divided equally ($N/3$) among the three groups, giving greater relative weight to long traces (usually fewer but richer in information). Proportional allocation: the number of traces per bin (N_b) is proportional to the actual bin size in the original log ($N \cdot |L_b|/|L|$), faithfully preserving the statistical distribution of the raw data. For variant selection and coverage constraints within each bin, traces are not chosen randomly, but based on their variant $v(\sigma)$ (the ordered sequence of activities). To ensure that the sample is representative and not dominated by a few repetitive processes, three selection methods are used: Top-k: guarantees the inclusion of at least one trace for the k most frequent variants. Proportional: selection is directly proportional to the frequency of the variant $f(v)$. Top-k-then-freq: a hybrid strategy that ensures the presence of the main variants and then fills the rest of the quota proportionally. To prevent extremely common variants from saturating the sample and obscuring rarer behaviors, a coverage constraint is imposed:

$$|S_v| \leq \alpha \cdot N_b$$

In this formula, α is an adjustable threshold (between 0 and 1) that defines the maximum share of the sample that can be assigned to a single variant. This mechanism is essential for preserving the so-called ”long tail” of clinical processes, ensuring that deviations or less frequent paths remain visible in the reduced log.

4.4.2 Directly-Follows Graph (DFG) Coverage

After the initial stratified sampling phase, the resulting sample is further refined to ensure that data reduction does not compromise the process topology. The goal is to achieve adequate coverage of the Directly-Follows Graph (DFG), ensuring that key transitions between activities are satisfactorily represented. A Directly-Follows Graph (DFG) is a fundamental representation

in process mining where nodes correspond to activities and edges represent direct succession relationships. Formally, the graph is constructed by extracting all pairs of directly consecutive activities $(a_i \rightarrow a_j)$ present in the log traces. To measure how representative the sample S is of the complete log L, the coverage coefficient $cov(S)$ is used, defined as the ratio of the direct relationships present in the sample to the total relationships of the original log:

$$cov(S) = \frac{|DFG(S)|}{|DFG(L)|}$$

. To refine the sample, a logic inspired by the RemainderPlus algorithm is adopted. The process follows a specific procedural path: Gap Analysis: The direct relationships of the sample are compared with those of the total log to identify missing transitions. Iterative Inclusion: Additional traces containing relationships not yet represented are iteratively added. Stopping Criterion: Insertion continues until the coverage reaches a predefined threshold γ (e.g., $cov(S) \geq \gamma$) or until a fixed maximum number of additions is reached. In this context, the γ threshold allows for balancing data synthesis with clinical accuracy. A high threshold ensures that even rare transitions (which could represent complications or less frequent critical paths) are included in the final model, preventing data reduction from ”hiding” process anomalies critical for medical validation.

4.4.3 Reduction Evaluation

The evaluation of the reduced sample is based on the analysis of structural and behavioral properties, with the aim of verifying that the reduction does not compromise the informative quality of the original log. To obtain a multidimensional evaluation, different categories of complexity metrics are adopted that go beyond the simple counting of events. The integration of heterogeneous metrics allows for a multidimensional assessment of log complexity. Specifically, metrics from different categories were considered:

- Quantity metrics: these metrics reflect the overall scale of the log, without providing information about internal variety:
  - Magnitude: Total number of events in the log, a direct measure of the size of the

dataset.

  - Support: Number of observed traces (cases). A reduction in support can result in a loss of behavioural coverage.

- Structure: These metrics describe the topological configuration of the process and the internal granularity of the traces:

  - Variety: Number of distinct activities present in the log. A good sample should maintain values similar to the full log, preserving semantic richness.
  - Level of detail: Average length of traces, i.e., average number of events per case. It measures how granular the paths are.

- Behaviour: these metrics relate to the variability of the observed behaviour, i.e., how structured or scattered the log is:

  - Deviation from randomness: Quantifies how much the observed sequences differ from random sequences generated on the same activities.

- Entropy Metrics: these metrics provide a probabilistic measure of process complexity, capturing behavioural distribution and uncertainty:

  - Variant Entropy: Shannon entropy on the distribution of variants. Higher values indicate greater uncertainty and behavioural variety.

$$H_{variant} = -\sum_{v \in V} p(v) \log(p(v)) \tag{4.1}$$

  - Sequence Entropy: Entropy calculated on sequences of activities, with two forgetting modes:
    * Linear forgetting: weights decrease linearly along the sequence;
    * Exponential forgetting: weights decrease exponentially.

- Algorithmic Metrics: LZC captures a domain-independent concept of complexity, related to information redundancy:

  - Lempel-Ziv Complexity (LZC) [9]: measures the compressibility of activity sequences. A log with many repetitions will have a low LZC, while one with very heterogeneous sequences will be less compressible and therefore more complex. It is calculated on activity sequences by applying the classical parsing algorithm.

- Organisational Metrics (Pentland): these metrics originate from organisational studies and are applied to process mining to describe the complexity of coordination and routines:

  - Task Complexity [10]: complexity of an activity viewed as an isolated unit, based on the variety of contexts in which it is compared. It measures the diversity of contexts in which an activity appears.
  - Process Complexity [11]: complexity resulting from the sequential articulation of activities within the overall process. It captures the articulation of sequential patterns.

In summary, the application of this heterogeneous set of metrics allows the validation of the effectiveness of the sampling strategy, ensuring a drastic reduction in data volume without sacrificing the dynamic fidelity and semantic richness of the original clinical processes.

4.5 Temporal Window Aggregation

Clinical logs frequently record events at a very fine temporal granularity. Activities that constitute a single intervention, such as continuous parameter monitoring or sequential data entry, are often represented as multiple events occurring within seconds. This results in excessive model complexity, which impedes interpretation of the logical information flow. The proposed aggregation method addresses this redundancy, simplifying the resulting graph while retaining essential semantic information. The technique relies on three key criteria for grouping events:

- Case ID: Aggregation occurs exclusively within the same process instance.

- Activity Label: Only events that share the same logical meaning (label) are grouped.
- Time Window: Defines the time boundary within which repetition is considered redundant.

Given a trace $t = \langle e_1, e_2, e_3, e_4, e_5 \rangle$ with labels $activity(t) = \langle a, b_1, b_2, b_3, d \rangle$, activities b_1 and b_2 are aggregated if the timestamp of b_2 falls within the range:

$$I = [b_1.timestamp, b_1.timestamp + window)$$

In this approach, only the first occurrence (b_1) is retained in the preprocessed log, while subsequent events that fall within the window are discarded. The refined log is used as input for the Directly-Follows Graph (DFG) algorithm, which operates by identifying ”directed succession” relationships:

- Nodes: Represent unique activities.
- Edges: If activity A is immediately followed by B in a trace, a directed edge $A \rightarrow B$ is created.
- Frequency: The model tracks how often an activity or transition appears, highlighting critical paths and bottlenecks.

To evaluate the impact of time window widths on model quality, the experimental phase examined a range of incremental intervals: from short-term (0–5 seconds, 1s steps) and medium-term (5–15 seconds, 5s steps) to long-term (up to 7 minutes, 15s steps), including a baseline version without aggregation for comparison.

To measure the effectiveness of the aggregation, metrics based on Token Replay are used, a technique that simulates the path of log traces within the discovered model. These metrics are: Fitness that measures how well the model is able to reproduce the behaviors recorded in the log, a high fitness indicates that the model explains the data well; Precision measures how restrictive the model is, i.e., whether it avoids allowing behaviors not present in the original log; and Fitting Traces measures the percentage of traces that can be reproduced by

Table 4.1: Example of extracting prefixes from a trace $t = (e_1, e_2, e_3, e_4, e_5)$ using a window size $w = 3$.

k	Prefix	Next activity
1	$\langle 0, 0, e_1 \rangle$	$\langle e_2 \rangle$
2	$\langle 0, e_1, e_2 \rangle$	$\langle e_3 \rangle$
3	$\langle e_1, e_2, e_3 \rangle$	$\langle e_4 \rangle$
4	$\langle e_2, e_3, e_4 \rangle$	$\langle e_5 \rangle$

the model from start to finish without errors. The underlying hypothesis of this work is that aggregation, by reducing noise and unnecessary repetitions, leads to increased precision and greater readability of the DFG, while maintaining optimal fitness levels for understanding the evidence-based process.

4.6 Extraction of Temporal Features for Next Activity Prediction

4.6.1 Prefixes Extraction

To predict the next activity, fixed-length prefixes are extracted from each trace using a sliding window size of $w = 3$. Consequently, each prefix is defined as a sequence of events $p = \langle e_{k-w}, \ldots, e_k \rangle$, where the objective is to predict the subsequent event $k+1$, with $k \in [1, M-1]$ and M representing the total number of events in the trace (length). In cases where the current position in the trace is smaller than the window size ($k < w$), a zero-padding technique is applied. This ensures that even the early stages of a process instance can be processed by the model using a consistent input shape, effectively capturing the immediate history preceding the activity to be predicted. An example is reported in Table 4.1 and the visual representation is reported in Figure 4.1

4.6.2 Temporal Features

Temporal features were extracted along three dimensions: inter-event latency, hour of the day, and day of the week. These variables were selected because previous studies in process

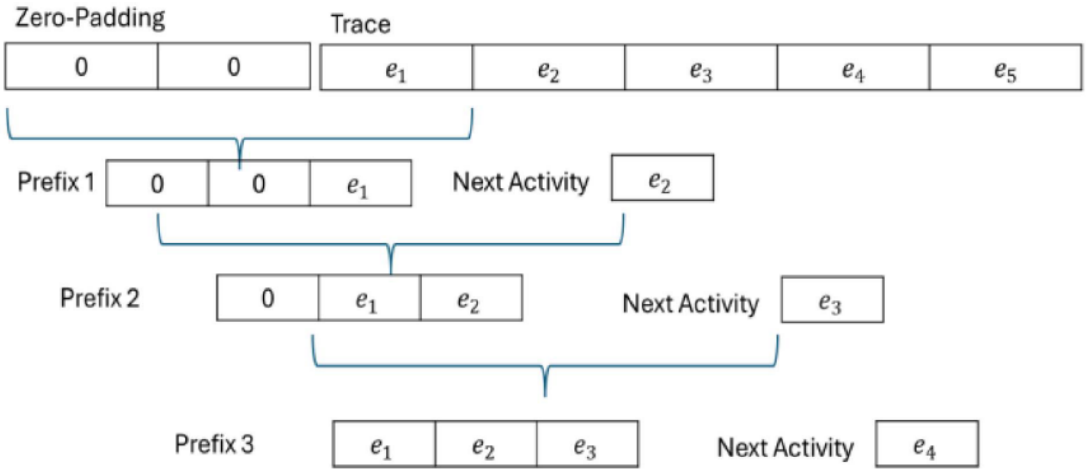

Figure 4.1: Representation of prefix extraction using a sliding window of size $w = 3$ with zero-padding.

mining have shown that temporal context significantly influences process behavior, especially in domains such as healthcare where operational workload and patient flow vary across daily and weekly cycles. The temporal information associated with each event is transformed into three primary dimensions to enrich the process representation and capture predictive patterns: TIMEDIFF, DAYTIME, and WEEKTIME[**aversano2025c**]. These features are derived from the events within the extracted prefixes, ensuring a sequence of temporal data that maintains consistent dimensionality.

- **TIMEDIFF.** This feature represents the time interval between consecutive events. For the initial event of a trace (e_1) or for events classified as zero-padding, the value is assigned as zero. In all other cases, the difference is computed by subtracting the timestamp of the preceding event from that of the current event.

  Formally, for a prefix p and a position j, the function is defined as:

$$timediff(p, j) = \begin{cases} 0 & \text{if } p[j] \text{ is padding} \\ 0 & \text{if } p[j] = e_1 \\ p[j].time - previous(p[j]).time & \text{otherwise} \end{cases}$$

- **DAYTIME.** This variable captures the hour of the day when the activity was recorded. The values belong to the range $[0, 23]$, allowing the model to identify daily routines or shift-related behavioral patterns.

- **WEEKTIME.** This feature identifies the day of the week, following the Italian encoding where the week begins on Monday (0) and ends on Sunday (6).

Once extracted, these temporal sequences are concatenated with the relevant prefixes. Furthermore, the activity labels themselves are processed using a Label Encoder, mapping each unique activity to an integer value from 0 to N. To evaluate the impact of this information, the experimentation follows an incremental approach. A *Base* configuration (using only activity labels) is compared against various combinations: *DIFF* + *WEEKTIME*, *DIFF* + *DAYTIME*, *DAYTIME* + *WEEKTIME*, and the complete configuration *DIFF* + *DAYTIME* + *WEEKTIME*. This structured comparison enables a precise assessment of how each temporal dimension contributes to improving the predictive performance of the model.

4.7 Predictive Modeling of Clinical Outcomes

This section describes the predictive modeling phase, which aims to predict the clinical outcomes of patients accessing the emergency department. The approach adopted is based on transforming the clinical event log into a structured patient-level dataset, followed by the application and comparison of various supervised classification algorithms. The methodology draws on and adapts the approach proposed in the MIMICEL study, applied to the emergency department context, with the aim of supporting clinical and organizational decision-making.

4.7.1 Patient-Level Dataset Construction

The emergency department event log is an event log extracted from the MIMICIV-ED dataset and describes the complete end-to-end process of a patient's journey in the emergency department (ED) and each event includes a case identifier (for each patient stay), a clinical-organizational activity, a timestamp, and static or dynamic attributes. This allows for the

analysis of existing patient flows, thereby improving the efficiency of processes within the emergency department. Each row in the CSV file represents the execution of an event during an ED stay, while each column corresponds to the specific attributes of that event. Initially, the dataset contains activities and several operations were conducted to transform the event log into a dataset suitable for machine learning techniques.

The information included in the dataset can be grouped into the following main categories:

- Patient demographics, such as gender and race;
- Method of arrival at the emergency department, distinguishing, for example, between independent access and ambulance transport;
- Vital signs, recorded upon admission and during the ED stay, including body temperature, heart rate, respiratory rate, oxygen saturation, and systolic and diastolic blood pressure;
- Clinical characteristics and reasons for the visit, including pain level, acuity, chief complaint, and derived semantic information (symptoms, body part involved, reason for admission);
- Treatment information is described indirectly through the presence and frequency of specific clinical activities.

The outcome of the ED stay is the target variable of the classification problem, this variable is represented by the disposition, which describes the patient's status at the end of the ED journey (e.g., discharge home, hospitalization, transfer, voluntary exit, or death). Given the strong heterogeneity and imbalance between classes, the dataset is appropriately cleaned, normalized, and, where necessary, balanced using oversampling techniques.

4.7.2 Preprocessing Operation

Initially, the clinical event log was transformed into a structured dataset, where each row represents a single patient and each column corresponds to a relevant feature collected during their length of stay to the emergency department. Unique case identifiers (caseIDs) were removed because they do not provide informative value for machine learning and may introduce bias

into the models. The preprocessing steps were taken to clean, normalize, and convert the data into a format suitable for training:

- Semantic Normalization of Clinical Text Variables: Based on the chief complaint column (the reason for the emergency department visit), three informative dimensions were extracted and standardized: the reported symptoms, the body part involved, and the primary clinical cause. These variables often contained heterogeneous values, including synonyms or alternative expressions for the same concept. A semantic mapping process was used to standardize the entries. For example: ”hypertension” and ”high blood pressure” $\rightarrow$ Hypertension, - ”hypotension” and ”low blood pressure” $\rightarrow$ Hypotension , ect. This process was extended to all textual columns to reduce dimension ality and improve semantic consistency across the dataset.

- Outlier Detection and Removal: Extreme and implausible values (e.g., heart rate or blood pressure readings exceeding 900) were identified and removed from the dataset.

- Temperature Normalization: Body temperature values were reported in both Fahrenheit and Celsius. All measurements were converted to Celsius, and unrealistic values were excluded.

- Pain Level Standardization: Subjective pain descriptions (e.g., ”mild,” ”moderate,” ”severe”,”unbearable”) were converted to numeric values on a scale from 0 to 10 through semantic mapping, allowing the pain level to be treated as a continuous numerical variable

- Handling Missing Values: A comprehensive analysis of missing or null values was conducted. Observations with missing values in critical variables were removed to ensure dataset quality and model robustness.

After thorough cleaning, transformation, and normalization, the final dataset was saved in a structured format (CSV), ready for training and evaluation of predictive models.

4.7.3 Process Mining Analysis

The process mining analysis was performed by performing analysis to extrapolate hidden information in the event log. The analysis was about discovering the Direct-Follow Graph (DFG), in order to highlight the relationship between activities, and the analysis the relation between the case's duration and case's number of activities. Direct-Follow Graph is a process model annotation system with the aims to represent processes as a graph using as relationship between activities the "direct follow" relation i.e. there is an arc between two activities a and b only if exists at least one trace where b occurs directly after a. Discovering DFG helps to understand the described behaviour by the event-log. Considering the variants -i.e. the different order of events that can be repeatedly recorded by an event-log - it was also discovered the DFG considering the top variants, in order to discover the most frequent trace recorded by the event-log. A further step was represented by understanding the relationship between the case's duration and number of activities. The aim was to understand if there was a change in the cases and to extract insight.

4.7.4 Classification

To classify the discharge status of patients from the emergency department (ED), a variety of classifiers were evaluated to identify the most effective model. Specifically, both tree-based models and ensemble methods were considered. Tree-based models operate by recursively splitting the data into smaller subsets, forming a hierarchical tree structure where nodes represent decision rules and branches represent outcomes. In contrast, ensemble methods combine multiple weak models to create a stronger classifier. The following classifiers were assessed:

- Decision Tree, used as a reference model for its interpretability;
- Random Forest and Extra Trees, which combine multiple decision trees to improve robustness and generalization;
- Boosting methods, including Gradient Boosting, XGBoost, CatBoost, and AdaBoost,

which build sequential models that can progressively correct the errors of previous classifiers.

These classifiers were selected for their diversity of approaches, all rooted in tree-based methods and ensemble learning. By comparing their performance, the objective was to determine the most suitable model for predicting patient discharge outcomes in the emergency department. For model evaluation, the Hold-Out validation method was used to split the dataset in 80% for training and 20% for testing. This partitioning allows the model to be evaluated on unseen data, providing a more reliable estimate of its generalization performance. The model's performance was assessed using the following metrics: Accuracy represents the percentage of correct predictions out of the total number of observations. It is useful in balanced datasets but may be misleading in cases of class imbalance; Precision indicates the proportion of true positive predictions among all instances predicted as positive, this metric is particularly relevant when the cost of false positives is high; Recall measures the proportion of actual positive cases correctly identified by the model, this is especially important when missing positive cases can have serious consequences (e.g., in medical diagnoses); F1-Score is the harmonic mean of precision and recall, providing a balanced metric even in imbalanced contexts, it is useful when a compromise between precision and recall is required.

4.8 Summary of the Methodology

The methodological framework presented in this chapter outlines an integrated analytical pipeline designed to systematically address the challenges posed by the stochastic, non-linear, and often fragmented nature of clinical processes. The primary objective of the entire architecture is the transformation of raw process data and heterogeneous event logs into a structured representation that is both computationally efficient and semantically dense, ensuring that clinically relevant information is preserved throughout the abstraction phases.

The logical path of the methodology is structured around four fundamental pillars:

- **Data Remediation and Integrity:** Building upon a rigorous preprocessing phase, the framework addresses log incompleteness through advanced repair techniques based

on *Snippet Extraction*. This approach enables the reconstruction of patient journeys as contextualized flows rather than isolated sequences. By probabilistically inferring missing activities, the framework restores the logical continuity essential for high-fidelity conformance checking.

- **Structural Complexity Optimization:** To manage the high dimensionality of clinical datasets—exemplified by the MIMIC-IV-ED repository—a stratified sampling strategy based on trace lengths and variants was implemented. Combined with *DFG Coverage* techniques, this ensures a reduced dataset that preserves the original process topology, maintaining the representation of rare transitions often indicative of clinical complications.

- **Abstraction and Noise Reduction:** The introduction of temporal window aggregation addresses the inherent complexity of "spaghetti" models in healthcare. By grouping micro-operational events into logical macro-phases, the framework increases the abstraction level and improves the readability of process graphs without compromising the model's fitness relative to empirical data.

- **Engineering the Predictive Potential:** Finally, the methodology converts the temporal dimension from a latent variable into an explicit predictor. Through the extraction of features such as TIMEDIFF, DAYTIME, and WEEKTIME, the log is transformed into a multidimensional input matrix. This ensures that subsequent machine learning algorithms can interpret the influence of circadian rhythms and operational latencies on hospital dynamics.

In conclusion, the synthesis of these procedures defines a robust and replicable research protocol. The internal consistency between the data cleaning, reduction, and enrichment phases ensures that the framework functions as an analytical ecosystem capable of faithfully reflecting the complexity of clinical practice. This solid methodological foundation serves as the essential prerequisite for the experimental phase, ensuring the reliability of predictions and the validity of the conclusions regarding workflow optimization and patient safety.

CHAPTER

5 Study Results

5.1 Data Acquisition & Standardization

This section presents the experimental results related to the first level of the proposed framework, addressing **RQ1** introduced in Chapter 3. The objective of this level is to ensure the acquisition of heterogeneous clinical data and their transformation into a standardized, consistent, and analysis-ready format suitable for process mining and predictive modelling.

The experimental evaluation focuses on real-world healthcare datasets characterized by heterogeneity in structure, semantics, and data quality. These datasets include clinical event logs extracted from different information systems, where inconsistencies, missing values, and unstructured textual attributes are common. Such characteristics make them representative of realistic healthcare environments and suitable for evaluating the robustness of the data acquisition and standardization layer.

The first step of the pipeline involves the ingestion of raw event data, including timestamps, activity labels, case identifiers, and additional clinical attributes. During this phase, structural inconsistencies such as duplicate events, invalid timestamps, and incomplete traces are identified and resolved. Events that do not conform to the minimum structural requirements for process analysis are filtered out, ensuring the integrity of the resulting event logs.

Subsequently, a semantic standardization process is applied to textual and categorical attributes. Clinical descriptions related to symptoms, causes of admission, body parts, and pain levels are normalized through controlled vocabularies and semantic mappings. This step sig-

nificantly reduces linguistic variability and dimensionality, enabling consistent interpretation of clinical information across different cases and time periods.

Numerical attributes, such as vital signs, are standardized by detecting and removing implausible values and by converting measurements to a common unit of reference where necessary. Missing values are handled according to attribute relevance, with incomplete records either excluded or corrected when reliable imputation is possible. As a result, the standardized datasets exhibit improved completeness and internal consistency.

The outcome of this first framework level is a set of high-quality, standardized event logs and tabular datasets that provide a reliable foundation for subsequent analytical layers. The experimental results demonstrate that the proposed data acquisition and standardization procedures effectively mitigate data quality issues, directly supporting the downstream process mining and predictive modelling tasks.

Overall, the results confirm that the framework successfully addresses **RQ1** by enabling the systematic transformation of heterogeneous and noisy healthcare data into structured and semantically consistent representations suitable for advanced analytics.

5.1.1 Experimental Design and Evaluation Strategy

This section introduces the experimental design adopted to evaluate the proposed framework and to answer the research questions defined in Chapter 3. The experiments are designed to assess the effectiveness of the framework across different analytical layers, focusing on data quality, process awareness, predictive performance, and adaptability in dynamic healthcare environments.

The evaluation strategy follows a multi-level approach, aligned with the layered structure of the framework. Each experimental setting targets a specific capability: data preparation and enrichment (RQ1), process reconstruction and enhancement (RQ2), temporal and behavioural analysis (RQ3), and predictive and adaptive modelling (RQ4). This design ensures that each component of the framework is validated both independently and as part of an integrated analytical pipeline.

Three real-world healthcare datasets were used in the experimental evaluation, each

characterized by different data quality issues, process complexity, and temporal dynamics. This choice allows the framework to be tested under heterogeneous conditions, reflecting realistic operational scenarios. In particular, the datasets exhibit incomplete logs, missing or noisy attributes, and evolving process behaviours, making them suitable for evaluating robustness and adaptability.

For the experiments involving predictive modelling, the evaluation follows a supervised learning setting. The available data are split into training and testing sets using a hold-out strategy, with 80% of the data used for training and 20% reserved for testing. This configuration allows for a reliable assessment of model generalization on unseen data. When class imbalance is present, oversampling techniques are applied exclusively to the training set to avoid information leakage.

Multiple performance metrics are adopted to ensure a comprehensive evaluation. Accuracy is used to measure overall predictive correctness, while precision, recall, and F1-score provide insight into class-specific performance, particularly in imbalanced scenarios. The use of multiple metrics is essential in healthcare contexts, where different types of errors may have different clinical implications.

To evaluate adaptability over time, the framework incorporates continuous monitoring mechanisms based on drift detection techniques. These mechanisms enable the identification of significant changes in data distributions or process behaviours, triggering model updates or retraining when necessary. Although the experiments are conducted offline, the evaluation simulates realistic deployment conditions by analysing performance stability across different temporal segments of the data.

Overall, the experimental design aims to validate not only the predictive accuracy of the proposed models, but also the framework's ability to support reliable, process-aware, and adaptive analytics in dynamic healthcare environments.

5.2 Data Quality & Repair

To address the second research question (**RQ2**: How is data quality ensured in clinical event logs?), this section evaluates the effectiveness of a *Supervised Machine Learning*-based approach for systematically repairing missing activity labels in clinical logs.

In real-world hospital settings, data integrity is frequently compromised by incomplete or incorrect records due to operational pressure, fragmented information systems, and the intrinsic complexity of clinical processes. Traditional *process discovery* techniques are often inadequate in the presence of such gaps, producing distorted or semantically inconsistent models.

The proposed approach addresses the data quality issue by transforming the repair of missing labels into a supervised predictive task, in which the missing activity is inferred by exploiting the sequential context of the event. This context is defined through:

- a *prefix*, consisting of previous activities;
- a *suffix*, consisting of subsequent activities.

This strategy preserves the semantic continuity of the process, avoiding the introduction of artificial noise into the repaired log.

5.2.1 Performance Analysis on Hospital Datasets

The experimental evaluation was conducted on two real clinical datasets, characterized by different levels of complexity:

- **Sepsis Registry**, representative of a highly variable clinical process driven by medical decisions;
- **Hospital Billing**, related to a highly structured hospital administrative process.

The results show that the quality of the repaired log depends directly on the amount of analyzed context, formalized through the *window size* (w), which ranges from 1 to 4 preceding and subsequent activities.

5.2.2 Effectiveness of Activity Reconstruction

Predictive Accuracy. In the **Sepsis Registry** dataset, the *Extreme Gradient Boosting* (XGBC) model achieved the best overall performance, reaching a maximum accuracy of **0.9171** with a contextual window size of $w = 4$, as reported in Table 5.1. This result highlights the model's ability to correctly reconstruct missing activities even in clinical processes characterized by high behavioral variability.

Table 5.1: Best results for the Sepsis Registry dataset

Window Size (w)	**Classifier**	**Accuracy**	**Precision**
1	GBC	0.8135	0.7157
2	XGBC	0.8890	0.7032
3	RFC	0.9065	0.7213
4	XGBC	**0.9171**	**0.7089**

In the **Hospital Billing** dataset, performance is significantly higher: accuracy consistently exceeds **0.99** across all considered splits, as shown in Table 5.2, indicating a near-perfect ability to identify missing activities. This can be attributed to the rigid and sequential nature of the administrative process.

Table 5.2: Best results for the Hospital Billing dataset

Window Size (w)	**Classifier**	**Accuracy**	**Precision**
1	XGBC	0.9908	0.8281
2	XGBC	0.9919	0.8741
3	XGBC	0.9927	0.8872
4	XGBC	**0.9930**	0.8472

Semantic Precision. Precision—a crucial metric for avoiding the generation of clinically implausible sequences—shows solid values in both datasets. Specifically:

- in the **Sepsis Registry**, precision remains stable above **0.70**;
- in the **Hospital Billing** dataset, values predominantly fluctuate between **0.85** and **0.90**.

These results confirm that the method does not merely fill in the gaps, but preserves the semantic coherence of the repaired process.

5.2.3 Impact of the Contextual Window Size

A key finding emerging from the analysis concerns the effect of the contextual window size. Increasing the number of preceding and subsequent activities considered leads to a systematic improvement in predictive performance.

Specifically, moving from $w = 1$ to $w = 4$:

- in the **Sepsis Registry**, an increase in accuracy of over 10 percentage points is observed;
- in the **Hospital Billing** dataset, accuracy—already high with smaller windows—reaches values close to perfection.

This trend is visually summarized in Figure 5.1, which illustrates the impact of the contextual window size on predictive accuracy for both datasets.

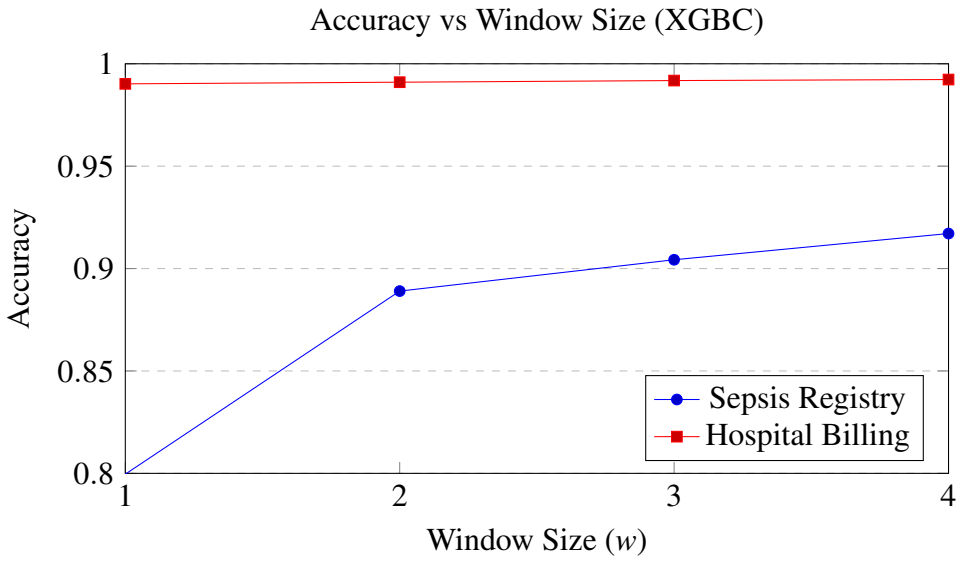

Figure 5.1: Impact of the contextual window size on reconstruction accuracy

This behavior demonstrates that clinical data quality is inherently **contextual**: a local view of the process proves insufficient for correctly reconstructing missing activities, whereas a wider bidirectional window allows for capturing the underlying clinical logic.

5.2.4 Summary and Conclusion for RQ2

The analysis of the experimental results on the **Sepsis Registry** and **Hospital Billing** datasets leads to three main findings:

- **Repair Reliability**: Supervised machine learning models ensure high levels of accuracy and precision, confirming their suitability for enhancing the quality of clinical logs.
- **Central Role of Context**: The inclusion of an extended bidirectional context significantly improves activity reconstruction by reducing semantic noise and systematic biases.
- **Superiority of XGBC**: The Extreme Gradient Boosting algorithm emerges as the top-performing model across nearly all experimental configurations, demonstrating particular effectiveness in modeling non-linear relationships between clinical activities.

In conclusion, these results demonstrate that the quality of clinical logs cannot be ensured through simple superficial data cleaning; instead, it requires an intelligent and contextual reconstruction of missing information.

The *Supervised Machine Learning* approach enables the transformation of fragmented and incomplete logs into coherent and semantically reliable representations of the actual process. This ensures that subsequent *Process Mining* phases reflect real clinical behavior rather than artifacts resulting from recording errors.

5.3 Model Discovery & Optimization

This section presents the experimental results obtained from the three pillars of the proposed framework: stratified sampling, temporal aggregation, and feature engineering. The goal is to demonstrate how these techniques combined ensure scalability, interpretability, and predictive robustness.

5.3.1 Results: Stratified Sampling for Log Reduction

The first study evaluated the impact of a stratified sampling strategy on process complexity metrics across three clinical datasets: **Sepsis Cases**, **MIMIC-ED**, and **COVID-19 CDSL**.

The results, summarized in Table 5.3, show that the sampling strategy successfully reduces the *Magnitude* (number of traces) and *Support* (number of events) by approximately 80

Table 5.3: Complexity Metrics: Original vs. Sampled Logs

Dataset	Type	Magnitude	Support	Variety	Granularity	Detail	Entropy
Sepsis	Original	1050	15214	16	14.49	0.0011	4.00
	Sampled	210	3045	16	14.50	0.0053	4.00
MIMIC-ED	Original	10000	56336	9	5.63	0.0001	3.17
	Sampled	2000	11271	9	5.64	0.0008	3.17
COVID-19	Original	2324	99872	512	42.97	0.0051	9.00
	Sampled	465	19974	512	42.95	0.0056	9.00

Discussion: The stability of *Activity Variety* and *Entropy* confirms that the sampling does not lose clinical information. The increase in the *Level of Detail* indicates that the sampled log is more ”dense” with information per event, making it more efficient for discovery algorithms.

5.3.2 Results: Model Simplification via Time-Window Aggregation

The second study addressed the complexity of discovered models (the ”spaghetti model” effect) using the **MIMICEL** dataset. The time-window aggregation technique was tested with windows of 1s, 60s, and 300s.

As shown in Table 5.4, the quality of the process model (Fitness and Precision) remains perfectly stable, even when the model structure is simplified by removing rapid-fire duplicate events.

Table 5.4: Quality Metrics for Time-Window Aggregation (MIMICEL)

Window	Trace Fitness	Log Fitness	Precision	F-Measure
Original	0.7335	0.6901	1.0000	0.8166
1s	0.7335	0.6901	1.0000	0.8166
60s	0.7335	0.6901	1.0000	0.8166
300s	0.7335	0.6901	1.0000	0.8166

The structural impact is visualized in the following chart, showing how the number of events is reduced without affecting model precision.

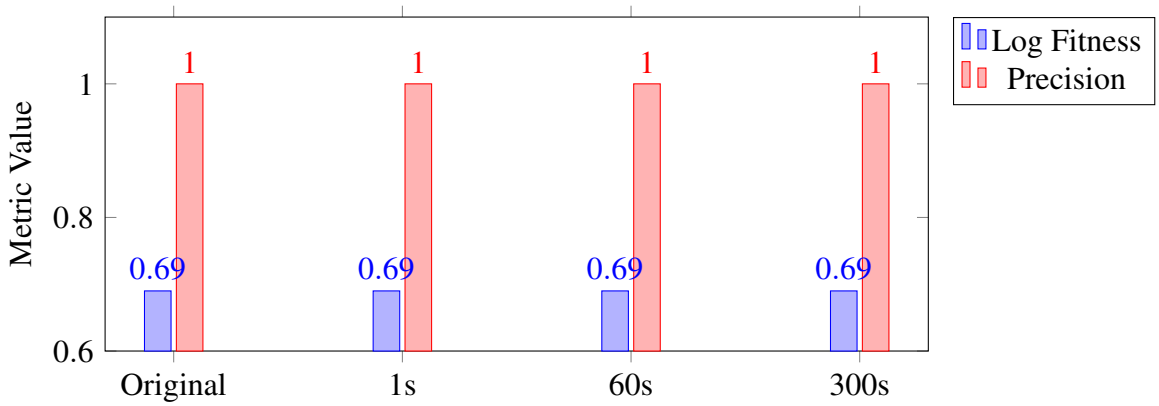

Figure 5.2: Stability of Model Quality across Aggregation Windows

5.3.3 Performance Gain from Temporal Features

The third dimension evaluated which temporal features are most effective for *Next Activity Prediction* using **BPIC 2012** and **BPIC 2017**. The comparison focused on the *TimeDiff* (time between events) versus baseline categorical features.

Table 5.5: Next Activity Prediction Results (Macro Average)

Dataset	Model	Accuracy	Precision	Recall	F1-Score
BPIC 2012	RF (Base)	0.7317	0.6558	0.6224	0.6224
	RF (**+TimeDiff**)	**0.7589**	**0.7258**	**0.6776**	**0.6776**
	XGBC (**+TimeDiff**)	0.7601	0.7248	0.6793	0.6793
BPIC 2017	RF (Base)	0.8038	0.7381	0.7229	0.7229
	RF (**+TimeDiff**)	**0.8242**	**0.7513**	**0.7466**	**0.7466**

Shapley Value Analysis: The study utilized SHAP values to interpret the models, revealing that *TimeDiff* is consistently the most influential feature for prediction, surpassing the contribution of activity labels themselves in certain process stages.

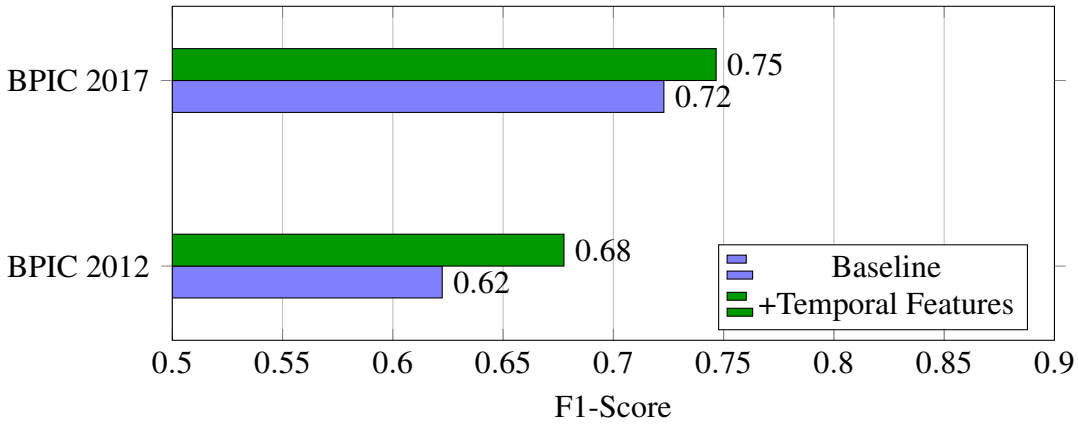

Figure 5.3: Comparison of F1-Score between baseline models and models enhanced with temporal feature engineering (TimeDiff). Results are averaged over 10-fold cross-validation.

5.3.4 Final Synthesis for RQ3

The combination of these results provides a comprehensive answer to RQ3. The **stratified sampling** ensures that the framework can process large-scale hospital data (reducing volume by 80%) without losing rare clinical paths. The **time-window aggregation** (optimally at 60s) simplifies the resulting process models, making them readable for medical staff without losing precision. Finally, the **temporal feature engineering** provides the necessary predictive accuracy (up to +5.5% F1-score) to support real-time clinical decisions.

5.4 Predictive & Adaptive Modelling

This section presents the results obtained from the predictive modelling layer of the framework, addressing **RQ4** and evaluating the effectiveness of process-aware and adaptive predictive models in emergency care. The analysis combines descriptive insights derived from process mining with quantitative performance metrics obtained from machine learning classifiers.

5.4.1 Process-Aware Descriptive Analysis

A preliminary analysis was conducted to characterize the behavioural dynamics of patient pathways within the emergency department. The Direct-Follow Graph (DFG) extracted from

the MIMICEL event log reveals a highly iterative and variable process structure. Core activities such as *Enter the ED*, *Triage in the ED*, and *Discharge from the ED* appear exactly once in every trace, defining the minimal clinical workflow.

In contrast, activities related to patient monitoring and treatment, including *Vital Sign Check*, *Medicine Dispensation*, and *Medicine Reconciliation*, exhibit multiple repetitions within the same case. This behaviour is reflected in the descriptive statistics, which show high variance in the frequency of these activities across traces. Such variability is clinically meaningful and supports the inclusion of frequency- and time-based process features in the predictive models.

The relationship between case duration and trace length was also investigated, revealing a moderate positive correlation ($\rho \approx 0.42$). Most cases are resolved within 24 hours and contain a limited number of events, while longer stays are associated with an increased number of activities and higher process complexity. These findings confirm that temporal and behavioural features provide valuable information for anticipating patient outcomes.

5.4.2 Baseline Predictive Performance

Initial experiments were conducted on the original, highly imbalanced target variable representing eight discharge outcomes. Without applying class balancing techniques, the best-performing models achieved accuracy values between 73% and 75%, with Gradient Boosting providing the highest performance among the evaluated classifiers.

However, the strong skewness of the discharge outcome distribution limited the reliability of these results, particularly for minority classes. This motivated the adoption of oversampling techniques to improve model robustness and generalization.

5.4.3 Predictive Results with Oversampling

After applying SMOTE to address class imbalance, a substantial improvement in predictive performance was observed across most classifiers. Table 5.6 reports the results obtained for the eight-class configuration. Ensemble tree-based models clearly outperform other approaches, with the Random Forest classifier achieving the highest accuracy (92.54%), along

with consistently high precision, recall, and F1-score values. Extra Trees exhibits comparable performance, confirming the effectiveness of randomized ensemble methods in capturing complex relationships between clinical variables and process-derived features.

Table 5.6: Classification results for the eight-class discharge outcome prediction (SMOTE applied).

Classifier	**Accuracy**	**Precision**	**Recall**	**F1-score**
Decision Tree	0.865	0.861	0.865	0.862
Random Forest	**0.925**	**0.924**	**0.925**	**0.924**
Extra Trees	0.924	0.922	0.924	0.923
Gradient Boosting	0.615	0.604	0.615	0.606
XGBoost	0.737	0.733	0.737	0.732
CatBoost	0.743	0.740	0.743	0.739
AdaBoost	0.466	0.454	0.466	0.457

To assess the impact of outcome aggregation, a second set of experiments was performed by reducing the original eight discharge categories to five macro-classes. The results, reported in Table 5.7, show a slight reduction in predictive performance. In this configuration, Extra Trees achieved the highest accuracy (88.63%), closely followed by Random Forest (88.58%). Although performance decreases marginally, the results remain robust and highlight a trade-off between predictive accuracy and interpretability.

Table 5.7: Classification results for the five-class discharge outcome prediction (SMOTE applied).

Classifier	**Accuracy**	**Precision**	**Recall**	**F1-score**
Decision Tree	0.818	0.815	0.819	0.816
Random Forest	0.886	0.885	0.886	0.885
Extra Trees	**0.886**	**0.885**	**0.887**	**0.885**
Gradient Boosting	0.709	0.706	0.709	0.706
XGBoost	0.774	0.772	0.775	0.772
CatBoost	0.777	0.775	0.777	0.775
AdaBoost	0.644	0.645	0.644	0.642

5.4.4 Comparative Analysis of Model Performance

Figure 5.4 summarizes the accuracy achieved by each classifier in both experimental settings. Ensemble tree-based models consistently dominate across configurations, while boosting-

based approaches show lower and more variable performance. This pattern suggests that models capable of handling high-dimensional and heterogeneous feature spaces are particularly suitable for emergency care prediction tasks.

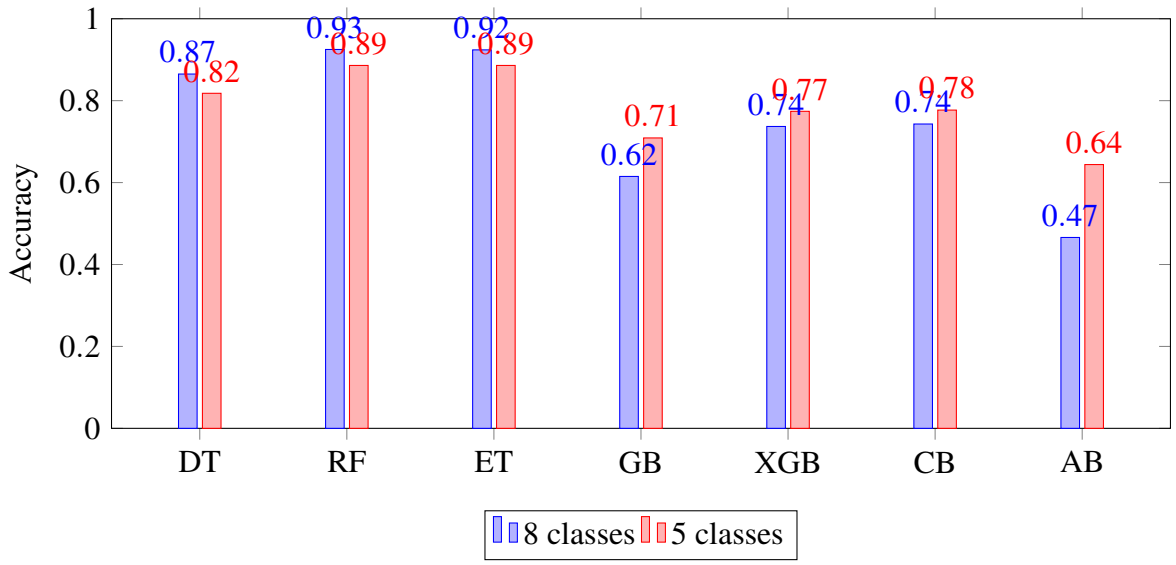

Figure 5.4: Comparison of classification accuracy across models for eight and five discharge outcome classes.

5.4.5 Discussion and Implications for Adaptive Modelling

The results demonstrate that predictive models enriched with process-derived and temporal features can accurately anticipate patient discharge outcomes in emergency care settings. The strong performance of ensemble tree-based classifiers confirms their robustness in handling dynamic, heterogeneous, and partially noisy healthcare data.

Importantly, the observed variability in patient trajectories and case durations reinforces the need for adaptive predictive systems. In such dynamic environments, continuous monitoring of data distributions and model performance is essential to prevent model decay. Within the proposed framework, this is achieved through drift detection mechanisms and periodic model retraining, ensuring sustained predictive reliability over time.

Overall, these findings provide empirical evidence in support of **RQ4**, showing that predictive models grounded in process mining and maintained through adaptive strategies can effectively support clinical decision-making in complex and evolving healthcare contexts.

5.5 Governance & Continuous Maintenance

This section discusses the results and implications related to the governance and long-term maintenance layer of the framework, addressing **RQ5** introduced in Chapter 3. While the previous sections focus on analytical performance, this level evaluates the framework's ability to ensure reliability, transparency, and sustainability over time in dynamic healthcare environments.

Healthcare data and processes are subject to continuous evolution due to changes in clinical protocols, organizational practices, and patient populations. As a consequence, both data quality and model performance may degrade over time if not properly monitored. The proposed framework explicitly integrates governance mechanisms to manage this evolution and to prevent uncontrolled model decay.

From a data governance perspective, continuous monitoring of data quality indicators is employed to detect anomalies such as shifts in attribute distributions, increases in missing values, or changes in semantic conventions. These indicators enable early identification of data acquisition or documentation issues, ensuring that the analytical pipeline remains grounded on reliable and standardized inputs, in line with the objectives of **RQ1**.

From a model governance perspective, predictive models are continuously evaluated using performance and drift indicators. Data drift detection techniques identify changes in feature distributions, while process-aware indicators monitor variations in control-flow behaviour, activity frequencies, and temporal characteristics. This dual monitoring strategy allows the framework to detect both explicit data changes and more subtle behavioural shifts in clinical workflows.

When governance mechanisms identify significant deviations, adaptive responses are triggered according to predefined policies. These responses range from increased monitoring and partial updates to full model retraining. All maintenance actions are logged and versioned, enabling traceability and accountability throughout the model lifecycle.

In accordance with **RQ5**, the framework supports a human-in-the-loop governance approach. Domain experts can review detected drifts, validate retraining decisions, and interpret

changes in clinical processes, ensuring that automated adaptations remain aligned with clinical expertise and organizational goals.

Overall, the results demonstrate that governance and continuous maintenance are essential to transform predictive analytics into a sustainable and trustworthy decision-support system. By explicitly addressing long-term control and adaptability, the framework successfully answers **RQ5**, ensuring that data-driven healthcare analytics remain effective beyond initial deployment.

5.6 Integrated Results Synthesis

This section provides an integrated synthesis of the experimental results presented in the previous sections. While Sections 5.1 to 5.5 focus on specific analytical layers of the framework, the goal of this section is to consolidate the empirical evidence and to highlight how the different components jointly contribute to addressing the research questions defined in Chapter 3.

Table 5.8 summarizes the main results obtained for each research question, linking analytical objectives, methodological choices, and experimental evidence. Rather than reporting individual performance metrics, the table emphasizes the role of each framework layer in supporting a coherent and end-to-end analytical pipeline for healthcare processes.

The results related to **RQ1** confirm that systematic data acquisition and standardization are essential prerequisites for reliable healthcare analytics. The transformation of heterogeneous clinical data into consistent and semantically normalized event logs enabled all subsequent process mining and predictive tasks, demonstrating that data quality directly impacts analytical robustness.

With respect to **RQ2** and **RQ3**, the experimental analysis shows that process mining and temporal feature extraction provide complementary perspectives on clinical pathways. Process discovery techniques reveal the structural organization of care pathways, while temporal and behavioural indicators capture execution dynamics and variability. Together, these results support a richer and more interpretable understanding of patient trajectories than what can be

Table 5.8: Summary of Experimental Results Across Research Questions

RQ	**Analytical Objective**	**Methods and Artifacts**	**Main Results and Evidence**
RQ1	Data acquisition and standardization	Data preprocessing, semantic normalization, quality control of event logs	Heterogeneous clinical data successfully transformed into structured and consistent event logs, enabling reliable downstream analysis
RQ2	Process reconstruction and pathway analysis	Process discovery algorithms, activity reconstruction, enhanced event logs	Care pathways reconstructed despite high variability; dominant patterns and deviations identified
RQ3	Temporal and behavioural analysis	Extraction of temporal metrics, activity frequencies, transition analysis	Temporal and behavioural features reveal significant differences in patient trajectories and process complexity
RQ4	Predictive and adaptive modelling	Supervised machine learning with process-derived and temporal features	Improved predictive performance compared to baseline models; robustness under class imbalance and temporal variation
RQ5	Governance and continuous maintenance	Drift detection, model monitoring, human-in-the-loop validation	Sustained model reliability through adaptive maintenance and transparent governance mechanisms

achieved through traditional data-centric analysis alone.

The findings related to **RQ4** demonstrate that predictive models benefit substantially from the integration of process-derived and temporal features. Compared to baseline models relying exclusively on clinical variables, the proposed approach achieves improved predictive performance and greater robustness to class imbalance and temporal variation. These results highlight the added value of process awareness in predictive healthcare analytics.

Finally, the results addressing **RQ5** show that governance and continuous maintenance mechanisms are fundamental to sustaining analytical performance over time. The combination

of data-level and process-level drift detection, together with adaptive retraining strategies and human-in-the-loop validation, enables controlled model evolution in dynamic environments.

Figure 5.5 provides a visual mapping between the layers of the proposed framework and the corresponding experimental outcomes. The figure illustrates how each analytical level produces specific artifacts and insights that feed into subsequent layers, ultimately supporting reliable, predictive, and adaptive decision support.

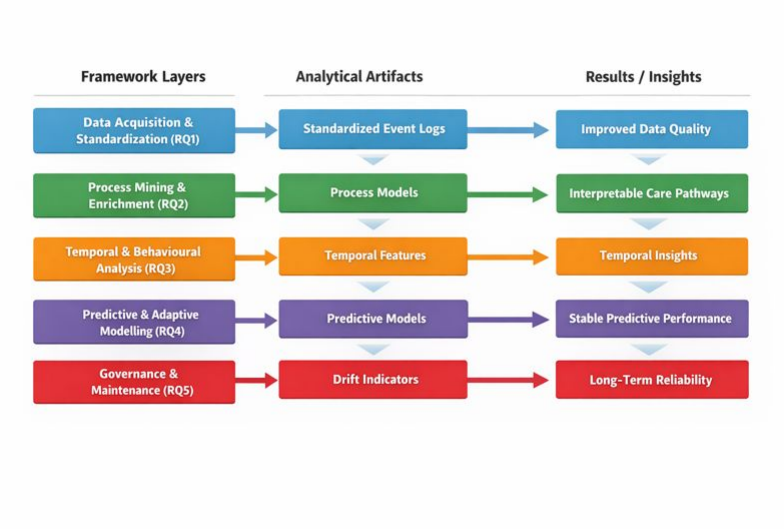

Figure 5.5: Mapping between the proposed framework layers and the experimental results obtained in Chapter 5.

Each analytical level contributes specific artifacts and insights, jointly supporting process-aware, predictive, and adaptive healthcare analytics.

Overall, the integrated synthesis demonstrates that the proposed framework should be interpreted as a cohesive system rather than as a collection of independent techniques. The experimental results confirm that the joint consideration of data quality, process dynamics, predictive modelling, and governance is essential for addressing the complexity of real-world healthcare environments.

CHAPTER

6 Conclusions

6.1 Conclusions and Insights

This thesis proposed a comprehensive, process-aware, and adaptive analytical framework for healthcare data analysis, with a particular focus on dynamic clinical environments such as emergency care. The framework integrates data acquisition, process mining, predictive modelling, and governance mechanisms into a unified pipeline. Through this multi-layered approach, the work addresses the challenges posed by data heterogeneity and temporal evolution that characterize real-world healthcare systems. The research successfully addressed five key research questions : Regarding RQ1, the framework demonstrated that systematic preprocessing and semantic normalization are essential to transform noisy clinical data into a robust foundation for analysis. For RQ2 and RQ3, process mining and temporal analysis proved effective in reconstructing care pathways and revealing execution dynamics that traditional data analysis fails to capture.The results for RQ4 confirmed that enriching machine learning models with process-derived features significantly improves predictive performance. Ensemble tree-based models, validated through 10-fold cross-validation, achieved an accuracy of 92.5%, demonstrating high stability and outperforming traditional baseline models. Finally, RQ5 was addressed by moving beyond static analytics. The proposed governance layer demonstrated that predictive pipelines can remain trustworthy through continuous monitoring and adaptive maintenance. Overall, this thesis contributes both conceptually and empirically to

the field of healthcare analytics by bridging process mining, machine learning, and governance principles. The proposed framework provides a structured and extensible solution for analyzing complex clinical processes while maintaining reliability over time. These contributions are particularly relevant for healthcare organizations seeking to leverage data-driven decision support in dynamic and high-stakes environments.

The synergy between process-aware feature engineering and automated drift detection transforms healthcare analytics from a descriptive exercise into a resilient clinical asset capable of adapting to evolving medical environments.

6.2 Threats to Validity and Limitations

Despite the promising results of this work, several limitations and threats to validity must be acknowledged to ensure a proper interpretation of the findings and guide future research.

Internal validity concerns the extent to which the observed results can be attributed to the proposed framework rather than confounding factors. Potential threats include the choices made during data preprocessing and feature engineering, which may influence the performance of process mining and predictive models. Alternative strategies in data cleaning, semantic normalization, or feature selection could yield different outcomes. Similarly, the use of oversampling techniques such as SMOTE improves model performance on minority classes but may introduce synthetic patterns that do not fully represent real clinical distributions. Moreover, all experiments were conducted in an offline evaluation setting, and although temporal segmentation and hold-out validation were applied, real-time deployment could expose models to dynamics not captured in retrospective analyses.

External validity and generalizability are limited by the scope of the datasets used, primarily derived from specific healthcare contexts. Healthcare processes vary significantly across institutions, regions, and specialties, meaning that findings may require adaptation to other settings. The absence of multi-institutional validation restricts the applicability of results, as differences in clinical protocols, documentation practices, and IT infrastructures may necessitate retraining and recalibration of predictive models. Additionally, the reliance on

retrospective data poses practical challenges for real-world deployment, including issues of data availability, latency, and integration with operational systems.

Construct and statistical validity concern the representation of theoretical concepts and the robustness of inferences. Complex clinical outcomes and process behaviors were simplified into discrete variables and aggregated features, which may overlook important contextual factors such as staffing levels, bed availability, or undocumented clinical activities. While multiple evaluation metrics were used to assess predictive performance, formal statistical significance testing was not applied across all experiments, so small differences in model performance should be interpreted cautiously.

Operational, socio-technical, and ethical limitations further constrain real-world applicability. Integrating the framework into clinical workflows requires careful consideration of interoperability, computational resources, and staff training, and resistance may arise due to trust or usability concerns. Ethical risks include the potential amplification of biases present in historical data, which necessitates continuous human-in-the-loop oversight. Implementing adaptive, real-time pipelines also demands scalable infrastructures and low-latency data processing to ensure reliability.

Finally, a distinctive methodological limitation is the ”digital shadow” bias inherent to process mining. Since the analysis relies exclusively on recorded events, numerous informal or verbal clinical interactions may be overlooked. This incompleteness can lead to reconstructed processes that misrepresent the true complexity of clinical workflows and potentially affect downstream analytical results.

6.3 Future Direction

Building upon these findings, several avenues for future work emerge. A primary objective is the rigorous application of the framework across diverse clinical institutions to strengthen its global robustness. Such multi-institutional validation is essential to evaluate the methodology's effectiveness across different hospital practices. The transition from batch-based monitoring toward real-time evaluation strategies represents a significant evolution. By implementing

live drift detection, the system could function as a dynamic decision support tool, minimizing the performance degradation window. Furthermore, deepening the integration of Explainable AI (XAI) is fundamental to fostering clinician trust and ensuring alignment with stringent regulatory requirements by providing clear ”why” explanations for model predictions. Finally, future efforts must prioritize causal inference techniques to distinguish between circumstantial correlations and actual drivers of clinical outcomes. This would transition the framework from predictive to truly prescriptive, offering evidence-based recommendations for process intervention.

The distinctive contribution of this thesis lies in the transition from static, data-centric analysis to an adaptive, process-aware, and governance-ready ecosystem. By bridging the gap between raw clinical events and sustained predictive reliability, the proposed framework provides a scalable foundation for AI that is not only accurate but also resilient, transparent, and ethically aligned with the future of digital healthcare.

List of Publications

The research activity carried out during the PhD program has resulted in the following scientific publications.

Journal Articles

- L. Aversano, M. Iammarino, A. Madau, G. Pirlo, G. Semeraro, *What Time Is It? Finding Which Temporal Features Is More Useful for Next Activity Prediction*, IEEE Open Journal of the Computer Society, vol. 6, 2025. DOI: 10.1109/OJCS.2024.3519815.

- L. Aversano, M. Iammarino, A. Madau, G. Pirlo, G. Semeraro, *Process mining applications in healthcare: a systematic literature review*, PeerJ Computer Science, 2025. DOI: 10.7717/peerj-cs.2613.

Conference Proceedings

- L. Aversano, M. Iammarino, A. Madau, D. Montano, C. Verdone, *Repairing Missing Activity Labels in Healthcare Process Logs: A Machine Learning Approach*, in *Innovation in Medicine and Healthcare (KES InMed)*, Springer Nature, Singapore, 2025. DOI: 10.1007/978-981-97-7498-2_9.

- A. Madau, G. Semeraro, *Process Mining and Machine Learning for Predicting Clinical Outcomes in Emergency Care: A Study on the MIMICEL Dataset*, Proceedings of the International Conference on Health Informatics (HEALTHINF), 2025, pp. 791–799. DOI: 10.5220/0013653500003967.

- L. Aversano, A. Madau, G. Semeraro, G. Verdone, *Effect of a Stratified Sampling Strategy on the Assessment of Process Complexity Metrics*, accepted for BAI'26 – The 2026 International Conference on Innovations in Computing Research. (DOI pending)
- A. Madau, G. Semeraro, G. Verdone, L. Aversano, *Analysis of Process Model Complexity through Time-Window Activity Aggregation*, accepted for BAI'26 – The 2026 International Conference on Innovations in Computing Research. (DOI pending)

Bibliography

[1] Wil van der Aalst, Ton Weijters, and Laura Maruster. “Workflow mining: discovering process models from event logs”. In: *IEEE Transactions on Knowledge and Data Engineering* 16.9 (2004), pp. 1128–1142.

[2] Wil M. P. van der Aalst. *Process Mining: Data Science in Action*. 2nd. Springer, 2016.

[3] Wil M. P. van der Aalst. *Process Mining: Data Science in Action*. 2nd. Berlin, Heidelberg: Springer, 2016.

[4] Wil M. P. van der Aalst. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer, 2011.

[5] Wil M. P. van der Aalst, Marlon H. Schonenberg, and Minseok Song. “Time prediction based on process mining”. In: *Information Systems* 36.2 (2011), pp. 450–475. DOI: 10.1016/j.is.2010.09.001.

[6] Michael Arias et al. “Mapping the Patient’s Journey in Healthcare Through Process Mining”. In: *International Journal of Environmental Research and Public Health* 17.18 (2020), p. 6586. DOI: 10.3390/ijerph17186586.

[7] L. Aversano et al. “Process mining applications in healthcare: a systematic literature review”. In: *PeerJ Computer Science* 11 (2025), e2613.

[8] L. Aversano et al. “Repairing missing activity labels in healthcare process logs: A machine learning approach”. In: *Innovation in Medicine and Healthcare (KES InMed 2024)*. Springer, 2025, pp. 91–101.

[9] L. Aversano et al. “What time is it? finding which temporal features are more useful for next activity prediction”. In: *IEEE Open Journal of the Computer Society* 6 (2025), pp. 261–271.

[10] R. P. J. C. Bose and W. M. P. van der Aalst. “Analysis of Patient Treatment Procedures: The BPI Challenge Case Study”. In: *Business Process Management Workshops (BPM 2011)*. Vol. 99. Lecture Notes in Business Information Processing. Springer, 2011, pp. 165–166.

[11] R. P. J. C. Bose and Wil M. P. van der Aalst. “Handling concept drift in process mining”. In: *Decision Support Systems* 56 (2013), pp. 83–99.

[12] L. Breiman. “Random forests”. In: *Machine Learning* 45.1 (2001), pp. 5–32.

[13] Manuel Camargo, Marlon Dumas, and Oscar González-Rojas. “Learning Accurate LSTM Models of Business Processes”. In: *Business Process Management*. Springer, 2019, pp. 286–302. DOI: 10.1007/978-3-030-26619-6_19.

[14] J. Carmona et al. *Conformance Checking: Foundations, Techniques and Applications to Business Process Management*. Springer, 2018.

[15] N. V. Chawla et al. “SMOTE: Synthetic minority over-sampling technique”. In: *Journal of Artificial Intelligence Research* 16 (2002), pp. 321–357.

[16] T. Chen and C. Guestrin. “XGBoost: A scalable tree boosting system”. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016, pp. 785–794.

[17] E. Choi et al. “RETAIN: An interpretable predictive model for healthcare using reverse time attention mechanism”. In: *arXiv preprint arXiv:1608.05745* (2017).

[18] Stefan Dakić, Olegas Vasilecas, et al. “Predictive process monitoring in healthcare: A case study using emergency department data”. In: *Artificial Intelligence in Medicine* 126 (2022), p. 102274.

[19] E. De Roock and N. Martin. “Process mining in healthcare – An updated perspective on the state of the art”. In: *Journal of Biomedical Informatics* 127 (2022), p. 103995.

[20] Emmelien De Roock and Niels Martin. “Process Mining in Healthcare – An Updated Perspective on the State of the Art”. In: *Journal of Biomedical Informatics* 127 (2022), p. 103995. DOI: 10.1016/j.jbi.2022.103995.

[21] A. V. Dorogush, V. Ershov, and A. Gulin. “CatBoost: gradient boosting with categorical features support”. In: *2018 IEEE International Conference on Data Mining (ICDM)*. 2018, pp. 660–669.

[22] Sebastian Dunzer et al. “Conformance checking: a state-of-the-art literature review”. In: *Proceedings of the 11th International Conference on Subject-Oriented Business Process Management (S-BPM ONE’19)*. New York, NY, USA: ACM, 2019, pp. 1–10. DOI: 10.1145/3329007.3329014.

[23] T. G. Erdogan and A. Tarhan. “Systematic mapping of process mining studies in healthcare”. In: *IEEE Access* 6 (2018), pp. 24543–24567.

[24] Andre Esteva et al. “A guide to deep learning in healthcare”. In: *Nature Medicine* 25 (2019), pp. 24–29. DOI: 10.1038/s41591-018-0316-z.

[25] Joerg Evermann, Jana-Rebecca Rehse, and Peter Fettke. “Predicting process behaviour using deep learning”. In: *Decision Support Systems* 100 (2017), pp. 129–140. DOI: 10.1016/j.dss.2017.04.003.

[26] A. L. Goldberger et al. “PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals”. In: *Circulation* 101.23 (2000), e215–e220.

[27] M. van der Heijden et al. “Process mining in emergency medicine: A systematic review”. In: *Journal of the American Medical Informatics Association* 27.9 (2020), pp. 1471–1483.

[28] H. Horita, Y. Kurihashi, and N. Miyamori. “Extraction of missing tendency using decision tree learning in business process event log”. In: *Data* 5.3 (2020).

[29] Peter B. Jensen, Lars Juhl Jensen, and Søren Brunak. “Mining electronic health records: towards better research applications and clinical care”. In: *Nature Reviews Genetics* 13.6 (2012), pp. 395–405. DOI: 10.1038/nrg3208.

[30] Alistair E. W. Johnson et al. “MIMIC-IV, a freely accessible electronic health record dataset”. In: *Scientific Data* 10.1 (2023), pp. 1–18.

[31] Ronny Mans, Hajo Schonenberg, and Minseok Song. *Process Mining in Healthcare: Evaluating and Exploiting Operational Healthcare Processes*. Springer, 2015.

[32] Travis B. Murdoch and Allan S. Detsky. “The inevitable application of big data to health care”. In: *JAMA* 309.13 (2013), pp. 1351–1352. DOI: 10.1001/jama.2013.393.

[33] Vincenzo Pasquadibisceglie et al. “A multi-view deep learning approach for predictive business process monitoring”. In: *Journal/Conference of Predictive BPM* (2022).

[34] Christoph Rinner and Erwin Helm. “Process mining in healthcare: A literature review”. In: *ACM Computing Surveys* 50.4 (2018), pp. 1–42.

[35] Enrique Rojas et al. “Process mining in healthcare: A literature review”. In: *Journal of Biomedical Informatics* 61 (2016), pp. 224–236.

[36] Suriadi Suriadi et al. “Event log imperfection patterns for process mining: Towards a systematic approach to cleaning event logs”. In: *Information Systems* 64 (2017), pp. 132–150.

[37] Niek Tax et al. “Predictive business process monitoring with LSTM neural networks”. In: *Information Systems* 80 (2019), pp. 1–17.

[38] Irene Teinemaa et al. “Outcome-oriented predictive process monitoring: Review and benchmark”. In: *ACM Transactions on Knowledge Discovery from Data* 13.2 (2019), pp. 1–57.

[39] E. J. Topol. *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. Basic Books, 2019.

[40] J. Wei et al. *MIMICEL: MIMIC-IV event log for emergency department (version 2.1.0)*. PhysioNet. Accessed: 2025-04-02. 2023.

[41] F. Xie, J. Zhou, J. W. Lee, et al. “Benchmarking emergency department prediction models with machine learning and public electronic health records”. In: *Scientific Data* 9.1 (2022).