

Article

PROTECTION: A BPMN-Based Data-Centric Process-Modeling-Managing-and-Mining Framework for Pandemic Prevention and Control

Alfredo Cuzzocrea ^{1,2,*}, Islam Belmerabet ¹, Carlo Combi ³, Enrico Franconi ⁴ and Paolo Terenziani ⁵¹ iDEA Lab, University of Calabria, 87036 Rende, Italy; islam.belmerabet@unical.it² Department of Computer Science, University of Paris City, 75006 Paris, France³ Department of Computer Science, University of Verona, 37129 Verona, Italy; carlo.combi@univr.it⁴ Faculty of Engineering, Free University of Bozen-Bolzano, 39100 Bolzano, Italy; franconi@inf.unibz.it⁵ DISIT Department, University of Piemonte Orientale, 15121 Alessandria, Italy; paolo.terenziani@unipmn.it

* Correspondence: alfredo.cuzzocrea@unical.it

† This research has been made in the context of the Excellence Chair in Big Data Management and Analytics at University of Paris City, Paris, France.

Abstract

The recent *COVID-19 pandemic outbreak* has demonstrated all the limitations of modern healthcare information systems in *preventing and controlling pandemics*, especially following an unexpected event. Existing approaches often fail to integrate *real-time data and adaptive learning mechanisms*, leading to inefficient response strategies and resource allocation challenges. To address this gap, in this paper, we propose PROTECTION, an innovative data-centric process-modeling-managing-and-mining framework for pandemic control and prevention that is based on the new paradigm that we name *Knowledge-, Decision- and Data-Intensive (KDDI) processes*. PROTECTION adopts *Business Process Model and Notation (BPMN)* as a standardized approach to model and manage complex healthcare workflows, enhancing interoperability and formal process representation. PROTECTION introduces a structured methodology that integrates *Big Data Analytics, Process Mining and Adaptive Learning Mechanisms* to dynamically update healthcare processes in response to evolving pandemic conditions. The framework enables *real-time process optimization, predictive analytics* for outbreak detection, and *automated decision support* for healthcare. Through case studies and experimental validation, we demonstrate how PROTECTION can effectively deal with the complex domain of pandemic control and prevention.

Keywords: process modeling; pandemic control and prevention; KDDI processes

Academic Editor: Domenico Ursino

Received: 4 July 2025

Revised: 8 September 2025

Accepted: 9 September 2025

Published: 22 September 2025

Citation: Cuzzocrea, A.; Belmerabet, I.; Combi, C.; Franconi, E.; Terenziani, P. PROTECTION: A BPMN-Based Data-Centric Process-Modeling-Managing-and-Mining Framework for Pandemic Prevention and Control. *Big Data Cogn. Comput.* **2025**, *9*, 241. <https://doi.org/10.3390/bdcc9090241>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Pandemics changed our daily life in many different aspects and, focusing on healthcare and clinical aspects, highlighted the need for managing newly branded *Knowledge-, Decision- and Data-Intensive (KDDI) processes* (e.g., [1–8]). We introduce this innovative definition to model and describe an approach designed to handle complex and evolving settings, such as healthcare systems impacted by pandemics. It uses an extensive amount of heterogeneous data (from sources like surveillance systems, IoT devices, and electronic health records), integrates expert knowledge (such as clinical guidelines), and helps make challenging decisions (in the face of uncertainty or changing conditions). The modeling, administration, and mining of healthcare systems enable real-time updates, predictive analytics, and adaptive responses to support pandemic prevention and control efforts. According to our vision,

KDDI processes are directly involved in both individual-level patient care, for example, we focus on managing *diagnostic* and *treatment workflows* for *swab-positive individuals*, and *population-level public health strategies*, such as the formulation and enforcement of policies aimed at mitigating disease transmission. By doing so, KDDI supports the dual objectives of preventing new outbreaks and controlling ongoing pandemic spread.

Beyond immediate, short-term interventions, KDDI methodologies also serve a critical role in long-term strategic planning and resilience building. The integration of *Information and Communication Technologies (ICT)* and *Artificial Intelligence (AI)* within KDDI frameworks enables robust mechanisms for data collection, analysis, storage, sharing, and visualization. These tools are instrumental in enhancing transparency and coordination among healthcare stakeholders.

Moreover, the recent global experience with COVID-19 has highlighted the necessity for sustained research and innovation in KDDI systems. This includes the development of new paradigms to better support *clinical and organizational decision-making*, with a focus on long-term capabilities for *disease surveillance*, *outbreak prediction*, *resource optimization*, and *patient-centered care planning*. Such efforts are pivotal in preparing healthcare systems to monitor, mitigate, and ultimately prevent future pandemic events.

In summary, the integration of KDDI processes into healthcare systems is not only a *reactive measure* to pandemic crises but also a proactive strategy aimed at transforming how health information is leveraged to inform policy, practice, and preparedness. In such context, *process modeling*, *management* and *mining* play a leading role, as to effectively support pandemic control policies at the large, with a special emphasis on the integration of these methodologies with the emerging *big data* trend, thus achieving the innovative definition of KDDI process modeling, management, and mining for pandemic scenarios. This, like in recent COVID-19 related studies (e.g., [3,4,9–11]).

While existing paradigms such as *Knowledge-Intensive BPM (KI-BPM)* [12] and *Data-Intensive BPM* [13] have addressed subsets of these challenges, they typically emphasize either knowledge modeling or data management in isolation. In contrast, we define KDDI processes as an explicit and systematic integration of three complementary perspectives: (i) *data-intensity*, through the ingestion and processing of heterogeneous real-time information such as *Electronic Health Records (EHRs)*, *IoT streams* and *epidemiological data*; (ii) *knowledge-intensity*, through the encoding and use of clinical guidelines, care pathways, and domain expertise; and (iii) *decision-intensity*, through the continuous support of adaptive policy-making under uncertainty. By formalizing this triadic combination, KDDI processes move beyond existing single-focus paradigms and provide a unifying framework for pandemic management where all three dimensions are simultaneously critical. Thus, KDDI does not propose an entirely new “*type of process*”, but rather a consolidated view that operationalizes the interplay of knowledge, decision, and data for resilience in pandemic response.

According to this last long-term perspective, we propose PROTECTION, a framework for supporting *data-centric process modeling*, *management* and *mining for pandemic prevention and control*. In this paper, we present a framework based on *Business Process Model and Notation (BPMN)*, which represents the most common standard for *modeling organizational processes* in a formal, yet accessible way. BPMN is particularly effective for capturing the *structure and dynamics* of KDDI processes, making available a notation that supports both human understanding and machine interpretability. Therefore, integrating BPMN into our framework contributes to the *transparency*, *adaptability*, and *interoperability* of process definitions.

PROTECTION, our proposed framework, builds upon these foundations. It provides support for *data-centric process modeling*, *management*, and *mining for pandemic prevention and*

control. PROTECTION focuses the attention on methodological issues in modeling, managing and mining healthcare/clinical KDDI processes for the management of worldwide pandemics. More into details, our proposed framework long-term aims are towards providing:

- (1) *clinical stakeholders* with a set of methodologies/tools to manage KDDI processes for the prevention and management of worldwide pandemics;
- (2) *healthcare decision-makers* with methodologies/tools for monitoring KDDI processes and resource consumption in their organizations, to control the care quality and the social impact of such pandemic-related processes;
- (3) *software designers* with a set of building-blocks and methodologies to support the efficient development of KDDI process systems devoted to the management of worldwide pandemics.

As regards the proper conceptual/software structure, PROTECTION is articulated into the following research assets/components:

- (1) *clinical and healthcare KDDI process modeling and management*, to represent knowledge of the target application scenario, plus its conceptual interconnections;
- (2) *clinical and healthcare KDDI process mining*, to both discover implicit processes (or process fragments) and to perform an “a posteriori” comparison between designed and actual processes;
- (3) *specific software architecture*, for: (i) modeling, managing and evaluating healthcare and clinical KDDI processes for preventing and managing pandemic events, and (ii) continuous KDDI process mining, to monitor actual processes and obtain useful feedback for improvement.

In addition to the framework long-term objectives, we emphasize the importance of *adaptive learning mechanisms* within PROTECTION. These mechanisms facilitate dynamic updates to healthcare and clinical processes based on real-time data and evolving pandemic conditions, ensuring a proactive and flexible response. By leveraging advanced machine learning techniques and predictive analytics (e.g., [14,15]), our system can identify emerging trends, potential hotspots, and bottlenecks in healthcare resource management. This allows for early intervention, which reduces the spread of infectious diseases and optimizes resource allocation.

The latter adaptive approach empowers clinical stakeholders to not only respond to immediate healthcare needs but also to forecast and prepare for future challenges, ensuring that preparedness strategies evolve in parallel with the progression of a pandemic. Furthermore, the integration of continuous monitoring, data aggregation, and real-time feedback loops ensures high flexibility and responsiveness in the decision-making processes (e.g., [16,17]).

The resulting dynamic, data-driven adjustment capability ensures healthcare system resilience, adaptability, and efficiency in managing current and future pandemic threats. By fostering continuous adaptation and learning, the system enhances the overall preparedness of healthcare infrastructures and improves outcomes during critical health crises.

1.1. Research Questions

The intrinsic goal of this paper consists in replying to the following research questions:

- **RQ1.** How can a data-centric approach enhance modeling, management, and mining of KDDI processes in pandemic prevention and control?

Data-Centric approach allows healthcare decision-makers to dynamically track pandemic evolution, patient care pathways, and resource consumption patterns. By integrating real-time data sources, such as Electronic Health Records, epidemiological surveillance

systems and IoT-enabled medical devices, PROTECTION ensures a continuous flow of critical healthcare information.

- **RQ2.** How can process mining techniques be leveraged to extract insights from healthcare and clinical data?

Process mining techniques are crucial for analyzing, optimizing, and enhancing healthcare processes during pandemics. The PROTECTION framework integrates process discovery, conformance checking and enhancement methods to monitor and refine pandemic-related healthcare operations.

- **RQ3.** What role do adaptive learning mechanisms play in dynamically updating healthcare processes?

Traditional pandemic response systems often rely on static healthcare guidelines that do not adapt quickly to evolving epidemiological patterns. Adaptive learning mechanisms, integrated within the PROTECTION framework, enable healthcare systems to dynamically update response strategies in real-time.

1.2. Paper Key Contributions

Following this main vision, this paper introduces and deeply discusses the framework PROTECTION. Rather than positioning PROTECTION as a completely new methodological framework, we emphasize its novelty as a *synthesis and application framework* that integrates and adapts existing concepts into a *coherent, data-driven process for pandemic prevention and control*. Specifically, we highlight how PROTECTION unifies knowledge discovery, decision-making, and data analytics perspectives into a single actionable platform, demonstrated through real-life case studies. In more detail, we describe anatomy and main functionalities of PROTECTION, along with several case studies showing how the proposed framework can effectively deal with the complex domain of pandemic control and prevention. In particular, our case studies also comprise several experimental parts where we clearly show how PROTECTION, via using *multidimensional big data analytics methodologies*, can really support actionable knowledge discovery and analytics, thus becoming an effective platform in the hands of healthcare decision-makers for pandemic prevention and control. Indeed, this main vision enforces the well-understood impact of *big data analytics technologies in healthcare*, as dictated by several recent studies in the area (e.g., [18,19]). Overall in this paper, we provide the following contributions:

- presenting PROTECTION as a *novel synthesis and application framework*, grounded in KDDI processes, to enhance pandemic prevention and control;
- providing a formalization of KDDI processes as a unifying paradigm that explicitly combines *knowledge-intensive*, *decision-intensive*, and *data-intensive* perspectives into a single methodological framework for complex healthcare scenarios;
- integrating adaptive learning mechanisms, big data analytics, and AI-driven predictive models to dynamically update healthcare processes based on real-time data and evolving pandemic conditions;
- employing process mining techniques to extract insights, compare designed and actual processes, and continuously improve healthcare operations for better pandemic response;
- validating PROTECTION through real-life applications, in order to demonstrate its effectiveness in managing pandemics by leveraging multidimensional big data analytics for decision-making and resource optimization.

1.3. Paper Organization

The remaining part of this paper is organized as follows. Section 2 focuses the attention on reviewing some relevant related work to our research. In Section 3, we present

the reference architecture of PROTECTION. Section 4 describes and details the anatomy and main functionalities of PROTECTION. After that, in Section 5, we provide several case studies demonstrating how PROTECTION effectively addresses the complexities of pandemic control and prevention. Section 6 provides a vertical scenario representative of the platform. Finally, Section 7 concludes the paper and provides future research directions for further expanding PROTECTION Framework.

2. Related Work

In this Section, we provide a comprehensive analysis of important and recent research proposals that are related to our work. Indeed, we can identify three relevant research areas that really influence our actions, namely: (i) *pandemic data source modeling*; (ii) *clinical guidelines and care pathways representation and management formalisms*; (iii) *process modeling and mining*.

2.1. Pandemic Data Source Modeling

How to model pandemic data sources? This challenging question can be investigated by carefully looking at the recent COVID-19 pandemic outbreak. Indeed, this critical event has attracted a lot of research in many intertwined fields, from *healthcare* and *medicine* to *bioinformatics*, from *data science* to *artificial intelligence*, from *risk analysis* to *multi-parameter optimization*, and so forth. Therefore, the issue of modeling and making publicly available COVID-19-related data and information (e.g., [20–22]) has been observed a great effort from the worldwide scientific community. Among these emerging kinds of data sources, which contain directions for modeling pandemic data with specific reference to COVID-19, we can identify the following ones.

First, the *European Centre for Disease Prevention and Control*, an agency of the European Union, provides a huge amount of *open healthcare data repositories* describing the worldwide history of this pandemic [23]. One of the main sources related to the evolution of the pandemic is the *COVID-19 Data Repository at Johns Hopkins University* [24]. Another example of a repository of multiple datasets related to healthcare and social COVID-related issues is [25]. As for the Italian context, the *Istituto Superiore di Sanità* provides information and also historical data about COVID-19 healthcare situation [22]. Second, *open clinical data repositories* are relevant to the scope of PROTECTION as well. Indeed, even though clinical datasets related to COVID-19 are complex to build and share for scientific purposes, some attempts have been made to allow scientists to analyze such data (e.g., [26–28]).

Further, since the treatment and prevention of COVID-19 patients received attention from worldwide healthcare institutions, which are providing a sort of continuously evolving recommendations, these can be freely interpreted as authoritative *clinical and healthcare guidelines*, being an invaluable source of novel research trends and conceptual blueprints for future research efforts.

The latter turns out to be effective under the form of procedures or technical guidance for different social, healthcare and clinical contexts (e.g., [29–31]). Finally, even *bibliographic repositories* are important sources of knowledge and information. Indeed, different publishers and health organizations launched different initiatives to achieve some shared effort to put at disposal the most recent scientific articles about COVID-19 (e.g., [21]), thus fostering research exchange and cooperation that, at the end of the day, result to be successful for enhancing the degree of innovation in the investigated research scenario.

2.2. Clinical Guidelines and Care Pathways

Clinical guidelines (GLs) consist of therapeutic and diagnostic recommendations encoding the “best practice” to care for specific patient categories. GLs are defined as “systemati-

cally developed statements to assist practitioner and patient decisions about appropriate health care in specific clinical circumstances". *Care pathways* (CPs) are instead defined as "structured multidisciplinary care plans which detail essential steps in the care of patients with a specific clinical problem" [32]. CPs are often the concrete application of GLs, where it is necessary to explicitly identify decision-based activities and all the complex clinical knowledge and data needed to suitably perform the planned activities. Both GLs and CPs are very relevant in PROTECTION, as they support knowledge modeling in the form of clinical and healthcare processes.

Several formalisms and tools have been proposed to represent, execute and verify GLs, often integrating formalized medical knowledge with data and workflow aspects, and supporting monitoring of GLs over time (e.g., [33]). A review of the state-of-the-art for these models for *Decision Support Systems* (DSS) has been published in [34,35]. When GLs are instantiated into a CP, their execution by various actors needs to be coordinated, and this may be performed both by *computerized guideline systems* and *Business Process Management* (BPM) *systems* (e.g., [36,37]).

2.3. Process Modeling and Mining

Clinical process management may also benefit from establishing a relationship with BPM [38,39], which can rely on a growing general interest and work on many proprietary and open-source tools. A plethora of data and information is generated within the execution of the clinical processes, thus fostering the adoption of BPM-like approaches to model and verify the observed behavior. The intrinsic complexity of the health field calls for models that reflect adaptivity to change, and that are able to deal with incomplete information, i.e., *models that enjoy flexibility*. At the same time, the involved entities are expected to behave in agreement with the specific medical/healthcare knowledge, regulations, norms, business rules, protocols and temporal constraints (e.g., [40]). Such GL systems (either BPM-based or not) require medical knowledge formalization, often relying on *Ontologies*. They have been extensively used in the medical domain since many years, but still deserve research efforts, in particular focusing on process-aware knowledge representation and on data-intensive process models (e.g., [41–44]).

Data from already-executed CPs would help to allow the discovery of "actual" processes, as well as their emerging correlations with healthcare and clinical data. Comparing designed processes and "actual" processes may help discover either errors in following a clinical guideline or new, partially unknown, *best practices* that could be suitably integrated into clinical guidelines/pathways. Recent approaches treating complex processes try to take advantage of distributed architectures, tackling the aspects of both mining new processes (e.g., [45]), complex multidimensional process mining (e.g., [46]), and monitoring the compliance of process executions (e.g., [47,48]).

Furthermore, we can acknowledge the amount of researchers' attention given for the integration of diverse data sources and methodologies to enhance the modeling of healthcare processes. Current research works (e.g., [49–51]) have emphasized the importance of synthesizing data from clinical records, real-time patient monitoring systems, social determinants of health, and external environmental factors to create more comprehensive and dynamic process models. This integrated approach not only improves the accuracy of pandemic predictions but also enables more effective resource allocation, optimized treatment strategies, and personalized care.

In addition, the incorporation of advanced artificial intelligence techniques, such as *Deep Learning* (DL) (e.g., [52,53]), *Natural Language Processing* (NLP) (e.g., [54,55]) and *Reinforcement Learning* (RL) (e.g., [56,57]), has shown promise in analyzing unstructured healthcare data, including medical notes, clinical research articles, and even patient-reported

outcomes, potentially leading to more *individualized treatment plans*. Additionally, the integration of *real-time data* from *wearable devices* and *IoT systems* offers a valuable layer of monitoring, enabling early detection of deteriorating health conditions or the emergence of new health threats.

By employing these technologies, healthcare systems can better understand complex patterns in patient care and outcomes, potentially leading to more individualized treatment plans. Additionally, the integration of real-time data from wearable devices and IoT systems offers a valuable layer of monitoring, enabling early detection of deteriorating health conditions or the emergence of new health threats.

Such intersection of AI, data integration, and healthcare process modeling holds significant potential for improving decision-making speed and accuracy, ultimately resulting in more proactive and effective pandemic responses. As the healthcare landscape continues to evolve, fostering cross-disciplinary collaboration and continuously refining these integrated models will be essential in building resilient, adaptable systems capable of effectively addressing both current and future global health challenges.

Table 1 summarizes the key existing approaches in pandemic management, comparing them based on their methodology. This comparison highlights the limitations of current systems and demonstrates how PROTECTION addresses these gaps in a number of ways as follows. (i) *Overcoming High Computational Resource Demands*, one of the primary limitations identified in the table under “Big Data Analytics for Pandemic Management” is the high computational resources required for analyzing large datasets. The PROTECTION framework mitigates this issue by leveraging cloud-based infrastructures and edge-computing architectures, which allow for distributed processing and more efficient resource utilization. This decentralization reduces the burden on central servers and facilitates real-time data processing, even in resource-constrained settings. Additionally, the use of lightweight machine learning algorithms within the framework reduces the computational intensity of predictive models, enabling faster processing with lower computational costs. (ii) *Addressing Data Privacy Concerns*, as data privacy and security are critical concerns in healthcare applications. While the PROTECTION framework does not explicitly focus on data privacy as the primary research area, it has been designed with privacy-preserving technologies such as federated learning and differential privacy. These technologies enable the analysis of healthcare data without exposing sensitive patient information, which is particularly important in the context of global health crises like pandemics. By utilizing federated learning, data can remain localized on patient devices, and only aggregated insights are shared, thus ensuring compliance with privacy regulations (e.g., GDPR, HIPAA). Moreover, the framework integrates *Secure Multi-Party Computation* (SMPC) protocols to protect data from potential breaches while still allowing for meaningful analysis of healthcare trends. (iii) *Real-Time Integration of Multidimensional Data*, the PROTECTION framework addresses the limitation of existing approaches, such as those found in process mining (e.g., [45–48]), by incorporating real-time, multidimensional data streams. Unlike retrospective process mining methods that primarily focus on historical data, PROTECTION enables dynamic process monitoring and adaptation. By continuously integrating real-time patient monitoring data, clinical records, and external environmental factors (e.g., weather patterns, social determinants of health), PROTECTION generates highly accurate and up-to-date process models. This allows healthcare systems to better anticipate and respond to rapidly changing pandemic conditions. (iv) *Scalability and Adaptivity*, the framework is designed to be scalable and adaptive, which is critical in the face of evolving pandemic scenarios. For instance, the PROTECTION framework incorporates adaptive decision-support mechanisms that can dynamically adjust to the changing nature of the healthcare environment, such as the appearance of new variants of a virus, changes in government policies, or fluctuations in

healthcare resources. This adaptability addresses the limitation in many existing pandemic management tools that struggle to respond to rapidly evolving circumstances. (v) *Incorporation of AI and Real-Time Decision-Making*, while existing systems (e.g., [49–57]) do not fully integrate AI-driven insights for decision-making, PROTECTION utilizes advanced AI techniques like DL, NLP, and RL for real-time predictions and personalized treatment plans. These AI models are trained on vast datasets, including *patient history*, *clinical guidelines*, and *environmental data*, allowing healthcare systems to quickly *adapt* to patient needs, *optimize resource allocation*, and *make informed decisions* during a pandemic.

Table 1. Comparison of Existing Approaches in Pandemic Management.

References	Approach	Methodology	Limitations
[23–28]	Pandemic Data Source Modeling	Uses healthcare records, epidemiological data, and open-access repositories	Lacks real-time updates and predictive analytics
[34–37]	Clinical Guidelines & Care Pathways	DSS and process modeling	Not adaptive to evolving pandemic scenarios
[38,39]	BPM for Healthcare	Workflow automation for clinical processes	Does not fully integrate AI-driven insights for decision-making
[45–48]	Process Mining in Healthcare	Extracts insights from clinical and operational data	Focuses on retrospective analysis, lacks predictive capabilities
[49–57]	Big Data Analytics for Pandemic Management	ML and predictive analytics	Requires high computational resources and data privacy concerns

Table 1 lists the approach-related drawbacks of the different pandemic management strategies. A data-centric, knowledge-driven, and decision-supportive approach based on KDDI procedures is integrated into the PROTECTION framework to fill these shortcomings. In contrast to many big data analytics solutions, which are limited to retrospective analysis and lack real-time flexibility, PROTECTION integrates adaptive learning processes and real-time process monitoring, enabling dynamic modifications to healthcare workflows based on real-time data. Furthermore, PROTECTION fills the gap between healthcare guidelines and responsive clinical operations by combining process mining and decision automation.

By utilizing knowledge-intensive structures like care pathways and ontologies, PROTECTION improves process models in contrast to conventional BPM systems, which frequently lack semantic depth and contextual awareness. We recognize that privacy and computational resource limitations, which are well-known problems in big data systems, are important concerns. While they fall outside the purview of PROTECTION at this time, our conclusions specifically name them as crucial areas for further investigation. All things considered, PROTECTION provides an expandable and integrated architecture that addresses the main issues with earlier pandemic prevention and control research.

3. The PROTECTION Reference Architecture

In this Section, we provide the reference architecture of PROTECTION, which has the final goal of capturing the many facets of pandemic prevention and control, as also demonstrated by the recent worldwide COVID-19 epidemic. The proposed architecture is modular in nature, and it unveils the complex interaction of process modeling, management methods, and data mining approaches in the context of treating such virulent viruses.

PROTECTION looks at pandemic events through a scientific lens, and it aims at uncovering the fundamental processes that regulate their *transmission patterns*, the efficacy of various intervention strategies, and the critical role of data-driven methods in shaping public health policy. Through rigorous analysis, this study not only elucidates the complexity inherent in pandemic management, but also emphasizes the importance of adaptable methods based on strong analytical frameworks.

Figure 1 shows the reference architecture of our proposed framework PROTECTION. Our reference architecture consists of several component layers, namely (i) *Big Data Sources* layer; (ii) *Big Data Storage* layer; (iii) *Process Modeling* layer; (iv) *KDDI Processes* layer; (v) *Big Data Analytics* layer. In the following, we describe these layers in detail.

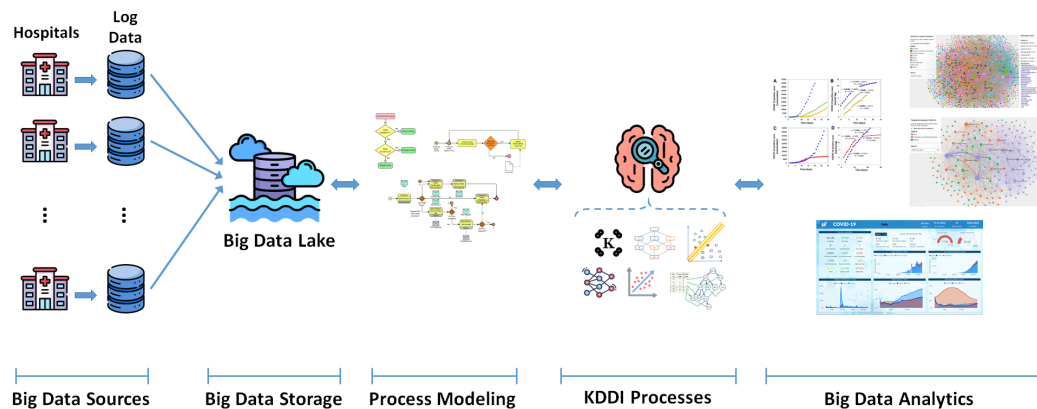


Figure 1. PROTECTION Reference Architecture.

Big Data Sources Layer. Healthcare data logs include valuable information such as patient demographics, clinical symptoms, laboratory findings, treatment procedures, and utilization trends. Using such comprehensive facts, we can build *complex models* that reflect the pandemic spatio-temporal development, evaluate the success of containment methods, and improve resource allocation in healthcare settings. Furthermore, using healthcare data logs allows for the inclusion of real-time information, permitting *dynamic-modeling techniques* that react to changing epidemiological patterns and healthcare demands. The rigorous examination of these massive data sources provides useful insights for optimizing pandemic response efforts and improving public health preparedness measures.

Big Data Storage Layer. *Cloud data lakes* provide scalable and cost-effective storage for pandemic-management data sources such as epidemiological surveillance, genomic sequences, healthcare records, and social media sentiment analysis. Using the flexibility and accessibility of Cloud infrastructures, we can seamlessly combine diverse information, allowing for *holistic modeling techniques* that reflect the complicated interaction of numerous factors influencing disease transmission and response tactics. By implementing Cloud data lakes, we want to demonstrate the usefulness of such storage solutions in empowering *data-driven insights*, therefore helping to the refinement of our process modeling framework and the optimization of pandemic mitigation efforts.

Process Modeling Layer. By taking advantages from advanced approaches like machine learning, statistical modeling, and NLP, we can extract *valuable insights* from various datasets stored in big data repositories. These datasets include epidemiological records, clinical data, genetic sequences, movement patterns, social media sentiment analysis, and other information, allowing for a more thorough knowledge of virus behavior and its implications for public health systems. This strategy not only allows for *real-time monitoring and prediction of disease patterns*, but it also enables evidence-based decision-making for successful pandemic management strategies.

KDDI Processes Layer. The incorporation of *knowledge-driven decision-making processes within data-intensive techniques applied on extensive Cloud data lakes* is a very effective approach that can be adopted in this layer of PROTECTION. By leveraging advanced methodologies such as machine learning algorithms, statistical modeling, and NLP, we can extract valuable insights from the diverse and voluminous big datasets stored within these repositories. These datasets encompass a wide array of information, including epidemiological records, clinical data, genomic sequences, socio-economic indicators, and mobility patterns. By harnessing the capabilities of big data analytics tools and techniques (e.g., [58–60]), coupled with knowledge-driven decision-making processes (e.g., [61]), we can also gain a comprehensive understanding of the pandemic dynamics (e.g., [62,63]). This facilitates informed decision-making in public health policy formulation, resource allocation, and intervention strategies aimed at mitigating the spread of pandemics and minimizing their impact on the society.

Big Data Analytics Layer. We can successfully manage and analyze massive amounts of heterogeneous big data stored in PROTECTION repositories by employing advanced approaches such as *distributed computing frameworks* (e.g., *Hadoop, Spark, Hive*, etc.), *scalable data processing engines*, and *Cloud-native analytics services*. Indeed, machine learning algorithms, DL models, and statistical techniques enable the extraction of significant insights from a wide range of datasets, including genomic sequences, clinical data, mobility patterns, sentiment analysis from social media platforms, and epidemiological records. These analytics tools and methodologies enable us to uncover hidden patterns, correlations, and helpful insights, which are crucial for driving evidence-based decision-making and establishing successful public health initiatives in response to the virulent epidemics. In particular, the strategy of PROTECTION consists in exploiting recent multidimensional big data analytics methodologies, given their proven effectiveness in several application scenarios, including healthcare analytics. Summarizing, these methodologies predicate the application of knowledge discovery techniques over *multidimensionally shaped big datasets*, in order to obtain all the benefits from powerful multidimensional modeling paradigms.

Building upon the detailed architecture of PROTECTION, we highlight the integration of real-time feedback mechanisms throughout each layer. This continuous feedback loop allows for iterative refinement of the pandemic management strategies in response to evolving data and circumstances. As new epidemiological data becomes available, it feeds directly into the Big Data Sources Layer, thus informing updated models and simulations within the Process Modeling Layer and KDDI Processes Layer. The described real-time data flow not only enhances the responsiveness of the system but also facilitates the rapid identification of emerging trends, potential risks, and areas requiring intervention.

By incorporating adaptive learning mechanisms (e.g., [64,65]) that continuously update based on the latest available data, the system can support proactive rather than reactive decision-making, leading to more effective containment strategies and resource allocation. Furthermore, this dynamic system enables the real-time evaluation of intervention measures, providing immediate insights into their effectiveness and allowing for prompt adjustments. The latter flexibility can enhance the integration of adaptive feedback mechanisms within the PROTECTION framework.

Specifically, integrating adaptive learning mechanisms can provide several key advantages to PROTECTION, such as:

- *Real-Time Process Optimization:* continuously refines healthcare workflows by analyzing real-time data and adjusting pandemic response strategies accordingly;
- *Enhanced Predictive Capabilities:* leveraging big data analytics to forecast emerging outbreaks, resource shortages, and potential healthcare bottlenecks;

- *Automated Decision Support*: assists healthcare decision-makers by dynamically updating guidelines and response measures based on evolving epidemiological trends;
- *Increased System Resilience*: enhances the adaptability of healthcare systems, ensuring they can respond effectively to new variants, changing public health conditions, and unforeseen crises.

4. The Emerging PROTECTION Methodology

The proposed framework PROTECTION is part of a long-term computer science and artificial intelligence project focusing on theoretical, methodological, and application-oriented aspects for the development of KDDI process systems able to deal with the complex domain of pandemic control and prevention. In this Section, we describe some important aspects of the emerging methodology induced by the overall PROTECTION proposal.

In order to support methodological issues in modeling, managing and mining issues, our proposed methodology is effective as it is supporting the pandemic control policies at the large, with a special emphasis on the integration of these methodologies with the emerging big data trend, thus achieving the innovative definition of so-called *data-centric process modeling, management and mining for pandemic scenarios*. As a proof of concept, PROTECTION targets the management of pandemics.

While a lot of attention arose on both healthcare and clinical data analysis and mining for pandemic management, little attention has been paid till now to some more long-term perspectives, mainly focusing on KDDI processes that use and generate such data. The main goal of PROTECTION is to propose a *methodological approach*, and some related software tools, to face future pandemics (and the continuation of the current one) by considering the healthcare and clinical processes enacted (and to enact) to fight the pandemics. Summarizing, from an attention to data we put the focus on KDDI processes, which have to be suitably designed and executed to take such critical pandemic under control, by a seamless integration of knowledge- decision- and data-related aspects.

The content of our proposed framework is drawn from *open-access repositories*, specific clinical and healthcare resources. Moreover, this content helps to serve as a reference for synthetic datasets. Also, we have used technical guidelines during pandemics for patients from USA, Europe, and the World Health Organization (WHO) [23,29,30]. History-oriented datasets from John Hopkins University [24] are also considered. Moreover, specific healthcare datasets were considered, related to the pharmacological monitoring of patients receiving monoclonal antibody therapies and the upcoming pharmacovigilance activities linked to pandemic-related vaccines. In terms of clinical datasets, we relied on certain clinical data repositories from the pandemic research database [28], which include electronic medical records of predominantly ambulatory patients.

For the objectives of our research project, we present a technique based on the generation of *synthetic datasets*. The synthetic datasets consist of multidimensional healthcare and epidemiological data, which were synthesized based on schema and attribute types observed in various public sources such as *EHR systems, pandemic surveillance systems, and process execution logs* (e.g., [24,28,30]). Each record in the synthetic datasets is modeled according to attributes such as (e.g., [31,35,37]): (i) *specific healthcare actions performed*; (ii) *classification of events* (e.g., *diagnosis, treatment, intervention*); (iii) *entities performing the action*, such as *healthcare providers*; (iv) *execution times*; (v) *anonymized patient identifiers*; (vi) *COVID-19 test type*; (vii) *prescribed treatments*; (viii) *recovery statuses* or further *medical actions* required. Thus, this makes available a realistic and *privacy-preserving* alternative for research in pandemic control.

As a consequence, real-time data streams from *IoT-enabled* medical devices are integrated into the datasets, facilitating adaptive learning mechanisms within our framework.

These streams enable *continuous monitoring* and *dynamic updates* to pandemic control strategies, supporting the development of predictive models (e.g., [42,45,61]). These structured synthetic datasets support *process mining*, *pattern discovery*, and *predictive analytics*, offering a *data-driven* method for supporting pandemic prevention and control efforts.

Furthermore, we have made an extraction of our synthetic datasets and the associated multidimensional models publicly available in an online open repository [66]. We were limited to this extraction due to project constraints. This will allow the scientific community to reproduce our study and adapt the datasets to other clinical or epidemiological use cases, thus contributing to broader research efforts in the field of healthcare and pandemic management.

Summarizing, the main axioms of the proposed PROTECTION framework are the following.

- *Modeling and Analyzing Healthcare KDDI Processes Dealing with the Management of Pandemics.* Such processes need to be designed and changed according to the possibly exponential diffusion of pandemics. They are characterized by many decision- and knowledge- intensive tasks. Here, integration with data (e.g., medical records, healthcare population data, and so on) and temporal constraints have to be considered. Simulation of such processes needs to be considered, to estimate feasibility, resource allocation, and so on. Different technical questions have to be addressed in this direction: how to represent medical knowledge of pandemic-related clinical guidelines? How do we merge and evaluate healthcare and clinical guidelines for pandemic prevention and patient management? How do we change healthcare processes according to the evolution of a pandemic? May we specialize healthcare pandemic control processes according to data coming from the pharmacovigilance for vaccines?
- *Pandemic-Related Process Mining, in order to Discover Process Models from Logs.* Whenever it is not possible to have log files to be analyzed in order to mine process models, the main idea is to consider both medical and healthcare records as an indirect kind of log, where therapeutic and specialized exams represent actions, main diagnoses represent (possibly) intermediate states of patients, and decisions for different allowed therapies/interventions/pathways represent knowledge-intensive decisional tasks. Here questions are like: May we discover some recurrent patterns of therapeutic actions/decisions not considered in the guidelines? Are the tasks recorded in medical records confirming the main indications of clinical and healthcare guidelines? Are there some suggestions in guidelines never considered in the medical records? May we suggest improvements for guidelines on the basis of the task patterns discovered from medical records? May we discover specific recurring care patterns for specific high-risk patients undergoing monoclonal antibody therapies?

In more detail, the main objective of our research is not to propose a novel methodology or framework but rather to leverage and apply an established approach, *Business Process Management* (e.g., [36,37]) within KDDI process paradigm, toward the domain of pandemic management. Our work focuses on demonstrating how BPM principles can be effectively integrated within the KDDI process to enhance *real-time decision-making*, *process optimization*, and *adaptive healthcare responses* in real-life applicative scenarios (i.e., *Pandemic scenarios*). Rather than defining entirely new process models or architectures, the manuscript illustrates how existing BPM techniques can be applied to real-life pandemic prevention and control scenarios, facilitating structured process modeling, management, and mining.

Reaching such goals would lead to significant advantages for the *National Healthcare System* (NHS) in promptly managing and preventing pandemic events. The progressive adoption of ICT techniques, in fact, can play a strategic role in the current rationalization process aimed at guaranteeing high-quality services, while reducing costs, even in a pan-

demographic event, where the management and prevention has to be enacted and monitored in a fast and dynamic way, to promptly react to diseases spreading with an exponential increase. Such a framework motivates the growing attention towards clinical and healthcare process definition and analysis.

PROTECTION pursues such goals through the development of several advanced and innovative research activities. In particular, process management in the clinical and healthcare domains is a significant topic, and we aim at bringing new challenges in the following research areas: ontological tools, languages based on different kinds of logics, data models and design tools for capturing events and temporal constraints, temporal extensions of GLs and CPs representation formalisms, constraint-based temporal reasoning, design-time and run-time GL verification, multidimensional analysis of healthcare processes, declarative and incremental process mining methods.

Specifically, PROTECTION introduces innovation in the areas of representation logic and ontological tools by structuring behavioral data in a multidimensional and semantically enriched manner. In terms of representation logic, PROTECTION incorporates hierarchical relationships, temporal dependencies, and attribute-based pivoting, enabling a flexible and expressive representation of behavioral responses to pandemic measures. This approach enhances the ability to reason over complex, evolving datasets and identify patterns that traditional flat representations might overlook.

It should be noted, here, that, even if the above-mentioned aspects are strictly related, so far, they have been considered in isolation and not yet applied cooperatively on the specific issue of managing worldwide pandemics. Starting from this limitation, PROTECTION aims at providing a set of methodologies and prototype software tools for the process-oriented prevention and management of worldwide pandemics.

The framework PROTECTION aims to integrate knowledge-driven decision-making processes into pandemic management by focusing on dynamic healthcare process modeling, management, and mining. The so-depicted methodology emphasizes data-centric process modeling, which integrates real-time healthcare data, clinical guidelines, and decision-making, allowing for adaptive and timely responses to evolving pandemic conditions. By incorporating big data techniques like predictive analytics and machine learning, the framework can detect emerging trends, predict resource needs, and provide early warnings.

Additionally, process mining techniques uncover inefficiencies or inconsistencies in the application of clinical guidelines, offering opportunities for process optimization. PROTECTION also promotes interdisciplinary collaboration, ensuring that solutions are robust, adaptable, and scalable across various pandemic scenarios.

For what concerns the *privacy* and *security* issues, these aspects are of critical importance, especially in the context of healthcare systems and private Cloud infrastructures. However, the focus of our research is specifically on process modeling, management, and mining within the BPM-KDDI framework for pandemic prevention and control. Our primary objectives are centered on demonstrating how Business Process Management can be effectively leveraged in real-life pandemic scenarios, along with addressing *security and privacy concerns*, *computational resources*, *real-time integration of multidimensional data*, *scalability and adaptivity*, and *the incorporation of AI and real-time decision-making*, which are relevant research domains in their own right. Nevertheless, we recognize the necessity of these aspects for real-life deployment and have explicitly integrated them as core components and mechanisms within our proposed framework for secure Cloud architectures, which further enhances the practical applicability of our approach.

Similarly, semantic aspects of data, such as *clinical data labeling*, are indeed crucial in the clinical domain, especially when dealing with large volumes of unstructured medical records and unlabeled databases. However, our research focuses on process modeling,

management, and mining within the BPM-KDDI framework, rather than on semantic data processing techniques. We acknowledge that clinical data labeling and semantic interoperability play a vital role in healthcare analytics. Nevertheless, incorporating these aspects requires specialized methodologies, such as NLP and *ontology-based approaches* (e.g., [67,68]), which differ in nature from the core objectives of our research. Therefore, while these aspects are highly relevant, their detailed investigation falls outside the scope of this paper and is left as future work.

Finally, the framework seeks to improve both the immediate response to health crises and the long-term management of healthcare processes, enhancing overall preparedness and resilience, which contributes to the broader advancement of healthcare process management in the face of future global health challenges.

We have also opted for the integration of adaptive learning mechanisms and real-time updates in the context of the BPMN process modeling within the PROTECTION framework. The latter addresses the need for dynamic updates to pandemic management strategies based on real-time data, and includes both specific mechanisms for adaptive learning and detailed explanations of the real-time scale and scope of process modifications.

Clarification of Adaptive Learning Mechanisms. To improve the flexibility and responsiveness of pandemic management, we integrate adaptive learning mechanisms into our BPMN-based process modeling. This Section now explicitly explains how real-time data analytics and continuous feedback from epidemiological trends drive dynamic updates to healthcare processes. Specifically, adaptive learning utilizes real-time event data such as *infection rates, hospital capacity, and emerging virus variants* to inform and adjust healthcare processes as conditions evolve. The adaptive learning mechanism continuously processes incoming data streams, allowing the system to recommend and apply modifications to BPMN models dynamically. This results in an up-to-date representation of healthcare workflows that adapts rapidly to the changing demands of pandemic control.

Real-Time Updates to BPMN Models. In order to further assess the feasibility of real-time updates, we provide a concrete example of how such updates occur within the framework. For instance, if a sudden surge in COVID-19 cases is detected in a region, the patient triage model could be dynamically updated to prioritize high-risk individuals or adjust treatment protocols in response to updated clinical guidelines. This update is enabled through the integration of process mining algorithms (e.g., *Alpha Miner* and *Inductive Miner*), which detect deviations from the predefined BPMN process model, and adaptive learning algorithms, which recommend modifications to the process flow based on new data inputs.

Types of Modifications and Scope of Updates. The types of modifications to the BPMN models are not restricted to simple rearrangements of tasks. Rather, the framework supports the addition of *new roles, tasks, or decision points* as necessary to address evolving healthcare requirements. For example, should a new patient care protocol become essential due to the emergence of a new variant of the virus, the framework allows for the integration of additional process steps or the introduction of specialized healthcare teams. Furthermore, process roles can be reassigned dynamically to match resource availability or evolving expertise requirements. This dynamic ability to modify the process models ensures that the healthcare system can respond quickly and efficiently to unforeseen changes in the pandemic landscape.

Triggering Events and Process Mining Algorithms. In this context, triggering events are essential for initiating updates to the process model. Events such as real-time infection rates, hospital occupancy, and patient outcomes serve as key signals for triggering updates to the healthcare processes. These events are processed by the process mining algorithms integrated into the framework, which include *Alpha Miner* and *Inductive Miner*. These algorithms *identify deviations from expected workflows, detect bottlenecks, and uncover ineffi-*

ciencies in the process flow. When such deviations are detected, the system initiates a review of the process model and proposes modifications based on the updated data and current operational needs.

Automatic Process Modifications and Clinical Guidelines Update. We further elaborate on the capability of the PROTECTION framework to autonomously update clinical guidelines based on real-time data analysis. For example, if the incoming data suggests a change in treatment protocols or prioritization, the system can autonomously propose updates to clinical guidelines, reflecting the latest available information. However, we clarify that although the process modifications can be automated, critical decisions, particularly those concerning patient care, will always involve some level of human oversight or validation to ensure that the system remains aligned with medical standards and ethical guidelines. This hybrid approach, where machine-based recommendations are supplemented by human decision-making, ensures that process optimizations remain grounded in clinical expertise.

Time Scales and Real-Time Scale of Updates. In response to concerns about the real-time scale of updates, we provide further clarification on the typical time frames involved in adaptive learning and real-time process updates. The frequency of process adjustments typically ranges from minutes to hours, depending on the nature and frequency of incoming data. For instance, when epidemiological models predict a sudden shift in transmission rates, corresponding updates to healthcare workflows (such as patient triage or resource allocation) can occur within a short time frame. This allows healthcare providers to rapidly adjust response strategies, thereby enhancing the effectiveness of pandemic control efforts.

Finally, the integration of adaptive learning mechanisms and the capability for real-time updates within the PROTECTION framework provides a significant advancement in pandemic management. These innovations enable *dynamic, data-driven decision-making* that can be promptly adjusted in response to emerging epidemiological trends, thus supporting the overall goal of improving pandemic preparedness and response strategies.

5. Pandemic Management and Control Measures Modeling

In this Section, we provide a few examples of pandemic management and control measures processes along with BPMN diagrams. These plans can be applied to various critical aspects of pandemic management, such as (i) *behavioral infection control* and (ii) *environmental pandemic control*. The strategic business objectives monitored by the platform include minimizing *ICU saturation*, optimizing *vaccination rollout*, and tracking the *emergence of new variants*, all of which play a key role in pandemic response. As a consequence, these objectives guide the design of the process models and analytics tools, ensuring that the data produced is actionable and aligned with overarching goals.

In addition to the BPMN diagrams, for each process model, we provide: (i) example datasets of data produced by process executions, which conforms to the datasets modeling described in Section 4; (ii) multidimensional big data analytics tools developed on top of these datasets, which are designed to support data-driven decision-making and optimize the control measures implemented. Specifically, indicators such as R_0 (*basic reproduction number*), *test positivity rate*, *ICU occupancy*, *vaccination coverage by demographic group*, and *average patient care delay* are carefully monitored, as these are critical for assessing the effectiveness of intervention strategies. However, due to project *data privacy agreements* and *non-divulgence constraints*, parts of the specific datasets and some process models used in this study are not made publicly available. These measures are put in place to ensure compliance with privacy regulations and protect sensitive information related to healthcare systems, patient data, and pandemic response activities. As a consequence, while the methodology, process models, and analytics tools are detailed in this paper, the actual data and some process-specific models remain confidential.

The business processes modeled include *diagnostic workflows*, *treatment administration*, *contact tracing*, *vaccination campaigns*, and *patient discharge planning*. Each of these processes represents a core area of pandemic response, and the BPMN diagrams are designed to clearly illustrate the flow of activities, decision points, and the interactions between different stakeholders. In fact, such models are particularly useful for optimizing resource allocation and streamlining operations, especially when faced with the challenge of large-scale pandemics.

To support the data modeling efforts, we use a multidimensional approach where facts and dimensions are defined clearly. Facts include daily case counts, hospitalization events, and intervention outcomes, providing the quantitative data necessary to evaluate the impact of various strategies. Dimensions, on the other hand, include factors such as *time*, *location*, *age group*, *healthcare facility*, and *intervention type*. These dimensions enable detailed analysis, allowing stakeholders to examine trends and patterns that are specific to different subsets of the population or geographies.

In this paper, we present a technique based on the integration of these facts and dimensions within a comprehensive data model. This method supports the creation of multidimensional data cubes that facilitate complex analytics, such as *trend analysis*, *predictive modeling*, and *scenario simulation*. Furthermore, these analytics tools can be tailored to support specific business objectives, such as *optimizing vaccination rollout* or *minimizing delays in patient care*, thereby enhancing decision-making capabilities.

The facts and dimensions used in our data modeling represent the building blocks for creating rich datasets that can be leveraged for advanced analytics. By interpreting these data in the context of business process models, the platform enables real-time monitoring of key indicators and provides actionable insights that can inform decision-making. As in many other cases, the integration of these models with big data tools is essential for adapting to the dynamic and unpredictable nature of pandemics, allowing for continuous updates to control strategies based on the most recent data.

Our approach also makes available a framework for ongoing learning, where adaptive strategies can be implemented as new information becomes available. The ability to track variant emergence and changes in epidemiological trends is particularly important in the context of a rapidly evolving pandemic. Therefore, by ensuring that the business processes are continuously updated and informed by the most recent data, the platform represents a powerful tool in the fight against pandemics.

However, it is important to note that the main issue in developing such systems lies in balancing the complexity of modeling and the practical need for real-time insights. From the above considerations, it follows that a methodical approach to process modeling and data analytics is more convenient than relying solely on real-time data collection without a structured framework. This allows for better preparedness and responsiveness, especially in addressing unexpected surges in cases or the appearance of new variants. By combining business process modeling with powerful analytics tools, the platform supports a dynamic and flexible approach to pandemic management.

5.1. Behavioral Infection Control

The process model *Behavioral Infection Control* is depicted in Figure 2, which illustrates the potential pathways of pandemic transmission in community settings and the corresponding behaviors capable of mitigating transmission. Isolation and social distancing measures, represented by the separation lane in the BPMN diagram, effectively block transmission by physically distancing infected individuals from others. However, the significant societal costs incurred, including economic, educational, and mental health impacts, highlight the complex challenges associated with adherence to these measures.

Moreover, without a viable vaccine, the relaxation of these measures may precipitate a resurgence of infections. Thus, widespread adherence to personal protective behaviors outlined in the model, such as proper cough etiquette, appropriate face mask usage, physical distancing, hand hygiene, disinfection of surfaces, and avoiding touching the face, is imperative. However, the effectiveness of these behaviors is contingent upon comprehensive guidance, training, and support to ensure their consistent implementation and thereby mitigate transmission effectively.

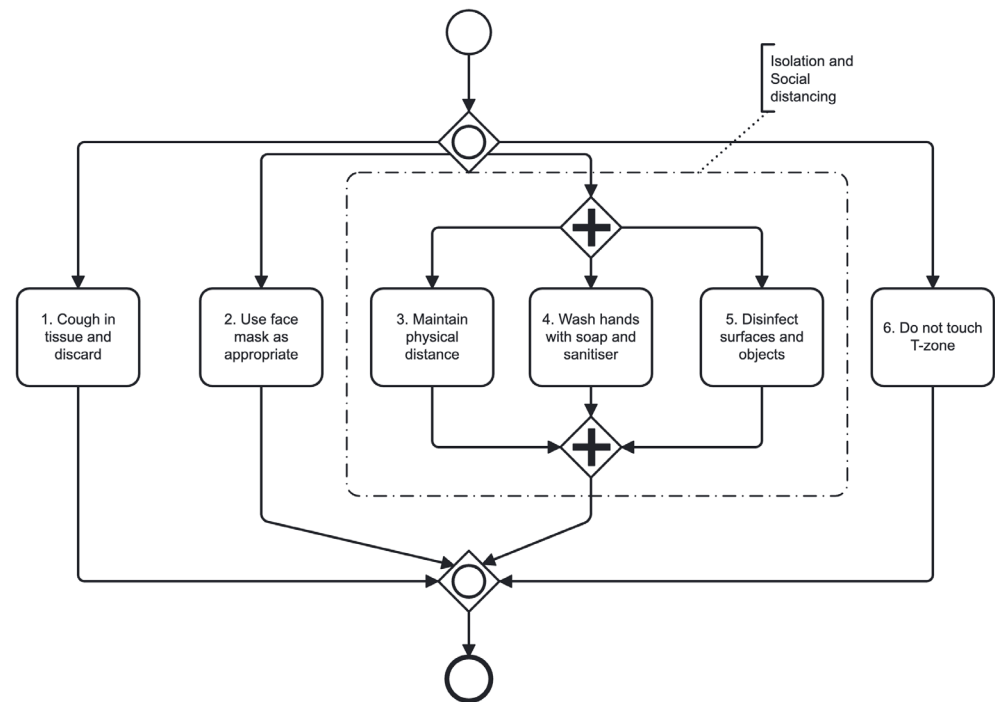


Figure 2. Behavioral Infection Control Process Model.

In this paper, we present a technique based on the extraction and modeling of a multidimensional dataset derived from log data, aimed at mining behavioral patterns related to pandemic preventive measures. A sample of the dataset for this running case study, named as *Behavioral Infection Control Dataset*, is shown in Figure 3. This dataset encompasses critical attributes necessary for analyzing human responses to infection control protocols within healthcare institutions. The most common fields include *Task_Name*, detailing the specific activity undertaken; *Event_Type*, delineating the nature of the action; *Originator*, identifying the individual initiating the action; *Timestamp*, providing temporal context; *Environment*, specifying the location or setting of the activity; *Surface*, describing the contact surface involved; *T-Zone*, indicating interactions with critical facial areas; and *ProtectiveDevice*, documenting the use of protective equipment.

From the above considerations, it follows that the kind of data structure used here is particularly suitable for modeling exposure dynamics and evaluating compliance. However, the main issue often lies in the perceived arbitrariness of certain column combinations. Therefore, clarification of column relationships was essential. In fact, fields such as *Surface*, *T-Zone*, and *ProtectiveDevice* are not arbitrarily assigned but are issued according to domain-driven association rules. For example, the task “Disinfect surfaces and objects” is not directly applied to the *Nose*; instead of implying such a literal interpretation, the *T-Zone* field represents areas of potential contamination or risk that may arise from neglecting this task.

In order to support our multidimensional big data analytics tasks, we modelled and created a suitable *multidimensional model*. Figure 4 presents the *Dimensional Fact Model* (DFM) of the *Behavioral Infection Control* dataset. The dimensions (red-colored in Figure 4) identified

in this schema are: (i) *Task_Name*; (ii) *Event_Type*; (iii) *Originator*; (iv) *Environment*; (v) *Surface*; and (vi) *ProtectiveDevice*. The selected measures are the count of distinct *Originator* and *ProtectiveDevice* values, respectively (blue-colored in Figure 4), which provides a robust quantitative basis for behavioral analysis. This multidimensional representation was generated by integrating behavioral event logs with predefined risk contexts (e.g., surface contacts and protective equipment usage), enabling the systematic association of tasks with exposure-relevant attributes. As a result, the dataset serves as a reliable analytical benchmark, facilitating both epidemiological assessments and behavioral pattern mining in infection control scenarios.

Task_Name	Event_Type	Originator	Timestamp	Environment	Surface	T-Zone	ProtectiveDevice
Wash hands with soap/sanitizer	Dismissed	P0128864	30/1/2024 2:00:30 PM	Waiting room	Table	Nose	Disinfectant gel
Wash hands with soap/sanitizer	Dismissed	P3544651	6/1/2024 2:00:06 PM	Meeting room	Chair	Mouth	Disinfectant gel
Wash hands with soap/sanitizer	Dismissed	P3135151	27/1/2024 2:00:27 PM	Waiting room	Table	Mouth	Mask
Wash hands with soap/sanitizer	Dismissed	P0648466	21/1/2024 2:00:21 PM	Meeting room	Desk	Nose	Gloves
Wash hands with soap/sanitizer	Completed	P4843135	6/1/2024 2:00:06 PM	Meeting room	Floor	Ears	Mask
Wash hands with soap/sanitizer	Completed	P0128864	3/1/2024 2:00:03 PM	Operating room	Chair	Mouth	Gloves
Wash hands with soap/sanitizer	Completed	P0648466	30/1/2024 2:00:30 PM	Meeting room	Equipment	Eyes	Gloves
Wash hands with soap/sanitizer	Dismissed	P0315468	30/1/2024 2:00:30 PM	Operating room	Chair	Ears	Tissue
Wash hands with soap/sanitizer	Completed	P8848646	21/1/2024 2:00:21 PM	Office	Desk	Mouth	Mask
Wash hands with soap/sanitizer	Dismissed	P0531548	30/1/2024 2:00:30 PM	Meeting room	Equipment	Nose	Tissue
Do not touch T-zone	Completed	P0531548	1/1/2024 2:00:01 PM	Operating room	Equipment	Nose	Mask
Disinfect surfaces and objects	Dismissed	P0654866	27/1/2024 2:00:27 PM	Meeting room	Table	Nose	Mask
Cough in tissue	Completed	P3544651	3/1/2024 2:00:03 PM	Office	Table	Ears	Disinfectant gel
Maintain physical distance	Completed	P4843135	12/1/2024 2:00:12 PM	Operating room	Table	Mouth	Gloves
Use face mask	Completed	P3544651	30/1/2024 2:00:30 PM	Office	Chair	Eyes	Disinfectant gel
Disinfect surfaces and objects	Completed	P0128864	21/1/2024 2:00:21 PM	Intensive care unit	Desk	Nose	Mask
Use face mask	Completed	P0128864	24/1/2024 2:00:12 PM	Waiting room	Floor	Ears	Tissue
Use face mask	Completed	P3544651	30/1/2024 2:00:30 PM	Operating room	Desk	Nose	Disinfectant gel
Disinfect surfaces and objects	Dismissed	P0128864	12/1/2024 2:00:12 PM	Meeting room	Floor	Nose	Tissue
Maintain physical distance	Dismissed	P0654866	24/1/2024 2:00:12 PM	Office	Chair	Nose	Tissue
Do not touch T-zone	Completed	P3135151	15/1/2024 2:00:15 PM	Meeting room	Chair	Nose	Gloves
Maintain physical distance	Completed	P0128864	6/1/2024 2:00:06 PM	Operating room	Chair	Ears	Mask
Disinfect surfaces and objects	Completed	P0648466	30/1/2024 2:00:30 PM	Meeting room	Desk	Eyes	Tissue

Figure 3. Sample from the Behavioral Infection Control Dataset.

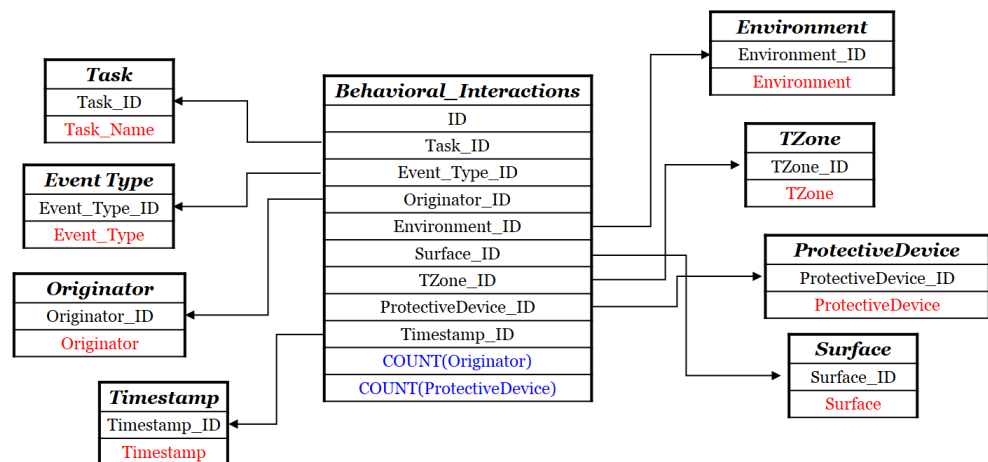


Figure 4. DFM over the Behavioral Infection Control Dataset.

We provide a fully annotated example, which represents an improvement over the previous version, in order to enhance transparency. The earlier sample has been replaced with a clearer, logically structured instance that explicitly traces how each value in the row is derived. By interpreting each attribute in relation to its contextual and procedural logic, the mapping is now more coherent and accessible. Obviously, these enhancements are aimed at facilitating more robust analytical interpretations of protective behavior in healthcare environments, as it strongly affects the efficacy of implemented measures.

The actions performed by an *Originator* (e.g., a healthcare worker or patient), such as “Cough in tissue” or “Maintain physical distance” are recorded based on log data collected from multiple sources within the healthcare environment. The collected data provide insights

into compliance with protective measures, behavioral trends, and intervention effectiveness. In our experimental setup, data collection follows a *batch processing approach* (e.g., [69,70]), where behavioral logs are aggregated over specific time intervals and later analyzed using our multidimensional framework. However, depending on the technological infrastructure, *real-time recording* can also be integrated when automated detection systems are available.

The dataset presented in Figure 3 has been extracted from log data recorded within a healthcare institution during the application of pandemic preventive measures. This dataset encompasses various relevant attributes and metrics for studying human reactions to protective protocols in actual healthcare environments. The data collection process relies on institutional monitoring systems, digital task management platforms, and automated logs of protective actions, ensuring accurate and continuous data recording. The dataset includes records of personnel activities, resource utilization, intervention tasks, and compliance behaviors, which provide insights into how individuals and teams adhere to pandemic control measures, optimize resource distribution, and respond to institutional protocols. The automatic logging of these activities enhances data reliability and reduces manual entry errors.

Beyond pandemic scenarios, such datasets can be used to monitor healthcare compliance, improve operational efficiency, and support future preparedness strategies. By analyzing behavioral patterns and intervention effectiveness over time, the framework remains sustainable and applicable even outside of pandemic periods, offering long-term benefits for healthcare management and policy-making.

Figure 5 shows an aggregation analysis bar plot representing the count of the healthcare institution staff members (attribute *Originator*) that have performed each type of prevention measure (attribute *Task_Name*).

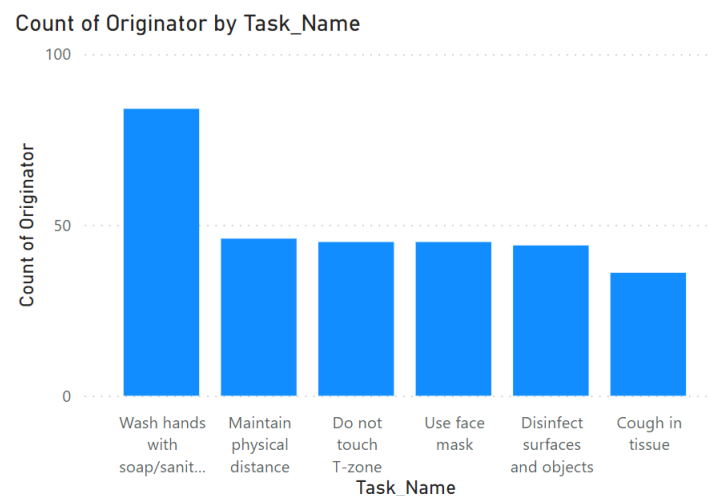


Figure 5. Aggregation Analysis of Healthcare Institution Staff Members that Performed Prevention Measures.

In Figure 6, we show a pivoting analysis by multi-attributes stacked column bar plot of the count of protective devices (attribute *Protected_Device*) used in each specific environment (attribute *Environment*) inside the healthcare institution.

Figure 7 shows the result of applying the clustering algorithm on the *Behavioral Infection Control* dataset presented in Figure 3. The clustering is performed on the prevention measures conducted by the healthcare institution staff members (attribute *Task_Name*) by the conductor of the measure (attribute *Originator*).

From this first experimental campaign, conducted on the Behavioral Infection Control Process Model (see Figure 2) and its corresponding dataset depicted in Figure 3, it follows

that employing *data sampling*, *aggregation analysis* of data, *pivoting analysis*, and *clustering* methodologies allows us to obtain significant insights into the dynamics of behavioral infection control during pandemics. Particularly, the model dataset provides essential multidimensional attributes that facilitate the understanding of waste management and medical interventions, contributing to more effective and timely decision-making processes.

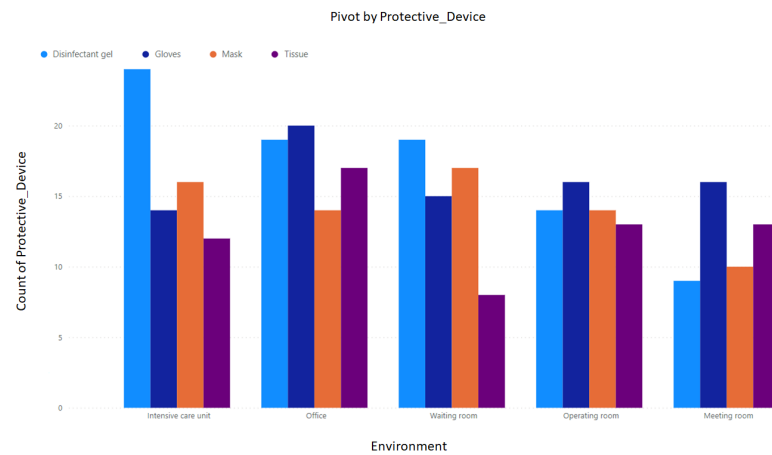


Figure 6. Pivoting Analysis of Protective Devices Used in Each Environment.

The adoption of data sampling, aggregation analysis, pivoting analysis, and clustering techniques within our proposed framework offers several advantages. Aggregation analysis allows the synthesis of detailed behavioral data into high-level indicators that facilitate the identification of key patterns and trends related to pandemic management. Pivoting analysis enables flexible exploration of the multidimensional dataset, revealing correlations and dependencies between behavioral attributes such as time, tasks, personnel, and resource utilization. Finally, clustering uncovers hidden patterns and groups of behaviors, supporting the identification of critical situations and the anticipation of future needs. The integration of these techniques into the proposed framework provides decision-makers with actionable insights and contributes to more effective and timely pandemic response strategies.

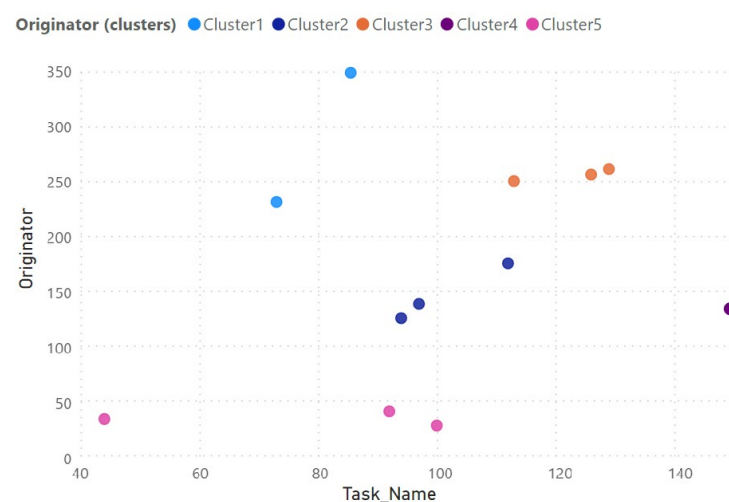


Figure 7. Clustering Results on the Behavioral Infection Control Dataset.

As a conclusion that can be drawn from this first experiment, the results highlight the effectiveness of the proposed framework in analyzing behavioral infection control measures during pandemics. The aggregation analysis quantifies critical behaviors such

as task execution and personnel actions, helping to identify trends and inefficiencies in pandemic-related waste management and medical interventions. Pivoting analysis provides a multidimensional view of the relationships between different attributes (e.g., time, personnel, and resource allocation), allowing for targeted decision-making. Additionally, clustering reveals hidden patterns in resource usage and task execution over time, offering valuable insights for anticipating future preventive actions. These findings confirm that our framework supports effective behavioral data analysis, enabling scalable, data-driven decision-making for pandemic response management.

These analytical tools are essential not only for Behavioral Infection Control but also for improving broader pandemic control efforts, such as *Environmental Pandemic Control*, *Pandemic Medication Planning*, *Pandemic Case Surveillance*, and *Pandemic Vaccination Planning*, where precise insights guide more effective planning and intervention strategies. These measures are described in detail in the following Sections.

5.2. Environmental Pandemic Control

The depicted business process model *Environmental Pandemic Control* of household waste management in Figure 8 delineates three distinct phases: storage, transport, and treatment, accommodating two categories of waste: pandemic virus-infected people waste and susceptible people waste. In the case of infected individuals, the storage phase involves specific tasks, including the segregation of waste from households with infected individuals or individuals in mandatory quarantine, utilization of bins with pedal-operated lids for mixed waste containment, and temporary suspension of source separation for recyclables and bio-waste. Subsequently, the transport phase entails ad hoc mixed waste transport services with segregated containers and increased collection frequency. Treatment primarily prioritizes mixed waste incineration or conventional methods such as *Mechanical-Biological Treatment* (MBTs) and *controlled landfill*. Conversely, for susceptible individuals, the storage phase maintains conventional waste containment practices with increased emphasis on proper bag sealing. Transport procedures involve heightened collection frequency for residual waste, with treatment strategies focusing on automated systems or preliminary storage for both recyclables and bio-waste. This comprehensive model provides a structured approach to household waste management tailored to different scenarios, thereby contributing to efficient resource utilization and public health protection amidst pandemic circumstances.

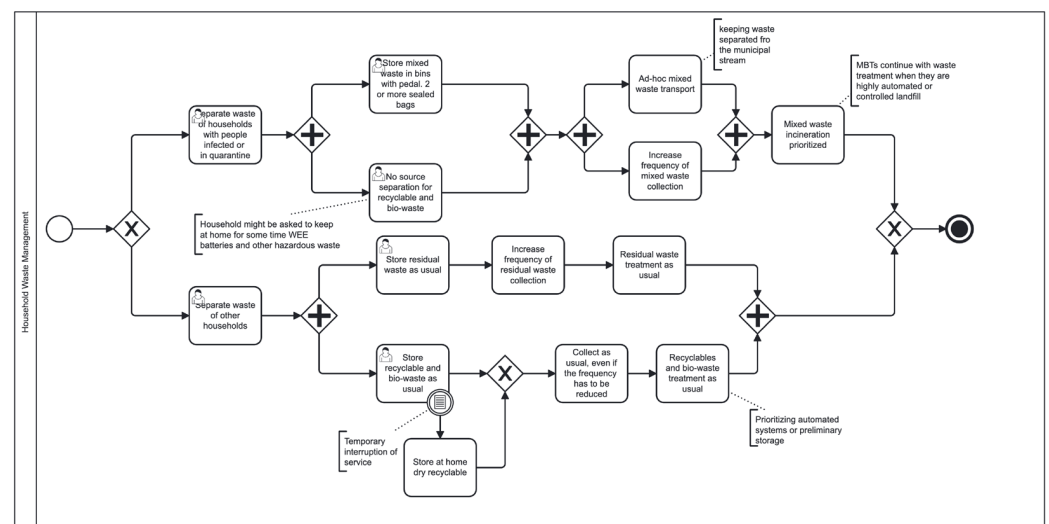


Figure 8. Environmental Pandemic Control Process Model.

In the case of behavioral pandemic fighting and preventive measures pertaining to municipal waste management, the extracted dataset encompasses critical multidimensional attributes/measures essential for understanding waste handling dynamics during health crises. Figure 9 presents *Dimensional Fact Model* of the *Environmental Pandemic Control* dataset. This diagram delineates how the dataset was constructed by linking recorded task events with environmental control variables related to waste management. Specifically, the model comprises the following key dimensions (red-colored in Figure 9): (i) *Task_Name*; (ii) *Event_Type*; (iii) *Originator*; (iv) *Waste*, *Separation*; (v) *Collection_Frequency*; (vi) *Waste_Treatment*. The selected measures are the count of distinct *Originator* and *Waste_Treatment* values, respectively (blue-colored in Figure 9). Each record captures a unique procedural instance, as identified by the *ID* and *Timestamp* fields. By formalizing the relationships among these variables, the schema clarifies the underlying data structure and supports a more systematic interpretation of environmental interventions. This structured representation enhances the dataset utility for downstream analyses, such as assessing adherence to waste handling protocols and identifying patterns in environmental control practices during pandemic conditions.

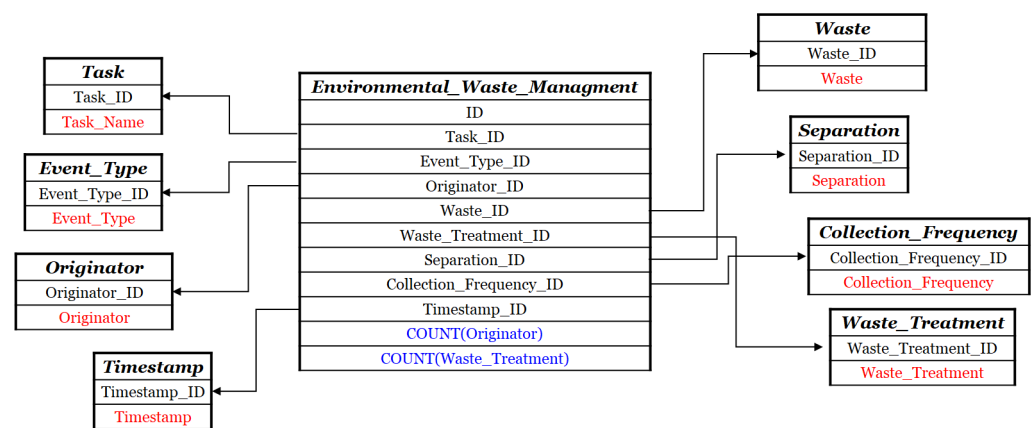


Figure 9. DFM over the Environmental Pandemic Control Dataset.

These attributes include *Task_Name*, delineating specific waste management activities; *Event_Type*, specifying the final state of each task execution (i.e., event); *Originator*, identifying the entity initiating the action; *Timestamp*, providing temporal context for each event; *Waste*, categorizing the type of waste involved; *Separation*, indicating the segregation process undertaken; *Collection_Frequency*, detailing the regularity of waste collection activities; and *Waste_Treatment*, elucidating the method employed for waste disposal or treatment. Such a dataset facilitates rigorous analysis of behavioral patterns in municipal waste management practices during pandemics, fostering insights into the effectiveness of preventive measures and their impact on waste handling strategies. In Figure 10, we provide a sample of *Environmental Pandemic Control* dataset that we have just described.

Figure 11 shows an aggregation analysis bar plot representing the count of the municipal waste management personnel (attribute *Originator*) that have performed each type of pandemic fighting and prevention measures (attribute *Task_Name*).

In Figure 12, we show a pivoting analysis by multi-attributes stacked column bar plot of the count of municipal waste treatments (attribute *Treatment*) performed in each specific preventive measure (attribute *Task_Name*) while the process of waste management. Similarly to the other cases, it should be noted here that powerful analytics methodologies allow us to really gain *knowledge insights* constituting the basis for superior KDDI processes.

Task_Name	Event_Type	Originator	Timestamp	Waste	Separation	Collection_Frequency	Waste_Treatment
Mixed waste incineration priori	Completed	P4843135	30/1/2024 2:00:30 PM	Infected mixed waste	Yes	1	Controlled landfill
Mixed waste incineration priori	Completed	TR8848646	24/1/2024 2:00:12 PM	Infected mixed waste	Yes	3	By operators
Mixed waste incineration priori	Dismissed	TS3135151	1/1/2024 2:00:01 PM	Infected mixed waste	No	2	Incineration
No source separation for recycl	Dismissed	TR8848646	12/1/2024 2:00:12 PM	Infected mixed waste	No	2	By operators
No source separation for recycl	Dismissed	TS0315468	12/1/2024 2:00:12 PM	Infected mixed waste	Yes	1	Incineration
Separate waste of households	Dismissed	P3544651	6/1/2024 2:00:06 PM	Infected mixed waste	No	3	By operators
Ad-hoc mixed waste transport	Dismissed	P3544651	6/1/2024 2:00:06 PM	Infected mixed waste	Yes	1	Incineration
Separate waste of households	Dismissed	TS0531548	1/1/2024 2:00:01 PM	Infected mixed waste	No	1	Controlled landfill
No source separation for recycl	Completed	TR0654866	18/1/2024 2:00:18 PM	Infected mixed waste	No	1	Controlled landfill
Separate waste of households	Dismissed	TS0531548	31/1/2024 2:00:31 PM	Infected mixed waste	No	3	Incineration
Separate waste of households	Completed	P3544651	24/1/2024 2:00:12 PM	Infected mixed waste	No	2	Controlled landfill
Separate waste of households	Completed	TR0654866	6/1/2024 2:00:06 PM	Infected mixed waste	No	3	Incineration
Separate waste of households	Completed	P0128864	27/1/2024 2:00:27 PM	Infected mixed waste	No	3	Controlled landfill
Mixed waste incineration priori	Dismissed	TS3135151	18/1/2024 2:00:18 PM	Infected mixed waste	Yes	3	Incineration
Store mixed waste in bins with	Dismissed	P4843135	15/1/2024 2:00:15 PM	Infected mixed waste	No	1	Incineration
Mixed waste incineration priori	Completed	TR0648466	31/1/2024 2:00:31 PM	Infected mixed waste	Yes	3	Incineration
Increase frequency of mixed wa	Completed	P0128864	27/1/2024 2:00:27 PM	Infected mixed waste	Yes	3	Controlled landfill
Ad-hoc mixed waste transport	Completed	P0128864	24/1/2024 2:00:12 PM	Infected mixed waste	No	1	Incineration
Separate waste of households	Completed	P0128864	31/1/2024 2:00:31 PM	Infected mixed waste	No	1	Incineration

Figure 10. Sample from the Environmental Pandemic Control Dataset.

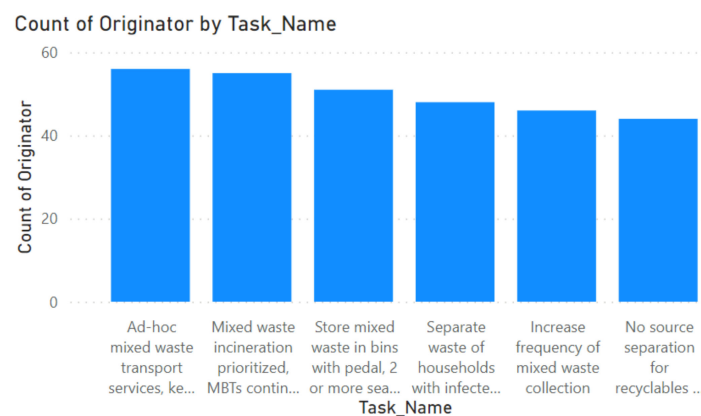


Figure 11. Aggregation Analysis of Municipal Waste Management Personnel that Performed Pandemic Fighting/Prevention Measures.

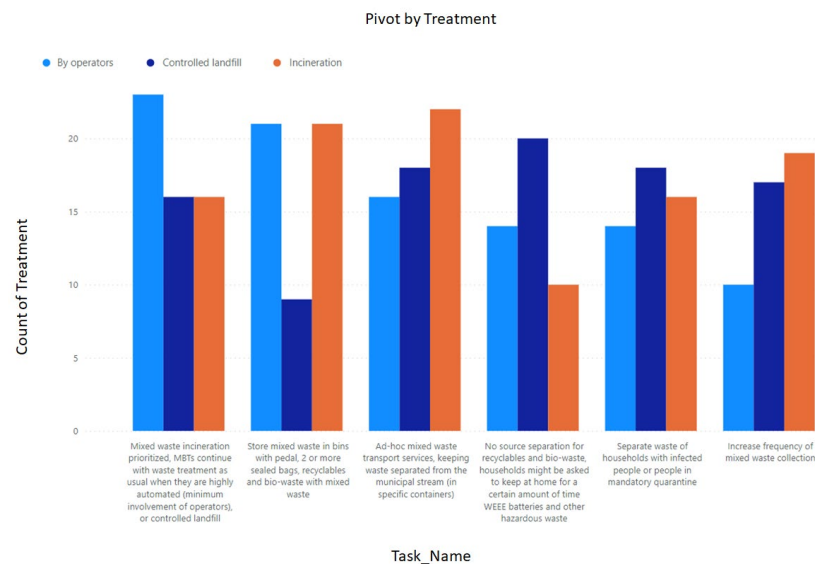


Figure 12. Pivoting Analysis of Municipal Waste Treatments Performed.

Figure 13 shows the result of applying the clustering algorithm on the *Environmental Pandemic Control* dataset presented in Figure 10. The clustering is performed at the time of execution of action (attribute *Timestamp*) by the executor from the municipal waste management personnel (attribute *Originator*).

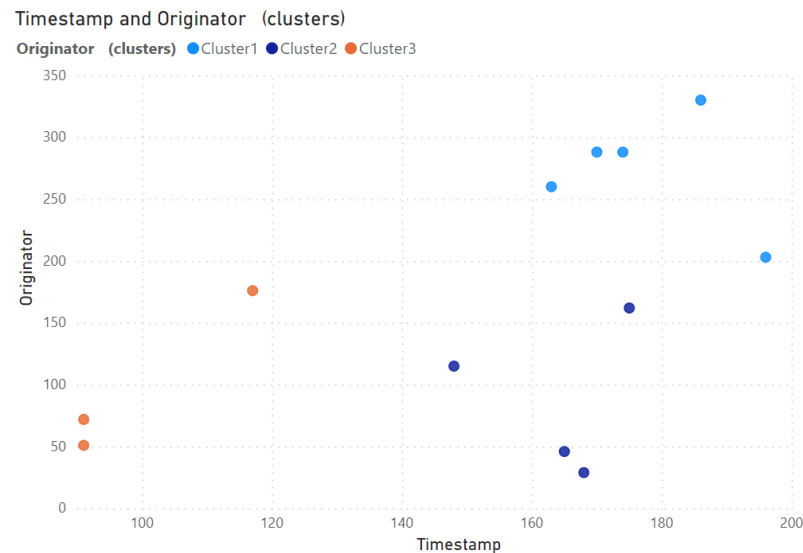


Figure 13. Clustering Results on the Environmental Pandemic Control Dataset.

Clustering analysis performed in this experimental campaign provides key insights into behavioral patterns in pandemic-related activities. The results reveal distinct groupings of task execution, personnel involvement, and resource utilization, allowing for the identification of patterns and anomalies in pandemic response strategies. This analysis highlights variations in workload distribution, enabling the detection of potential bottlenecks in medical interventions and waste management processes. Furthermore, clustering helps in predicting future trends by identifying recurring behavioral patterns over time, supporting proactive decision-making.

5.3. Pandemic Medication Planning

The process model *Pandemic Medication Planning* is depicted in Figure 14, which represents the process we proposed for the identification of patients testing positive by a particular virus pandemic. It is initiated by the reception in consultation on-site or teleconsultation of a patient presenting symptoms. A doctor or a health operator receives the patient and creates a new patient file for him if he is not already registered on the platform, otherwise, he analyzes his request.

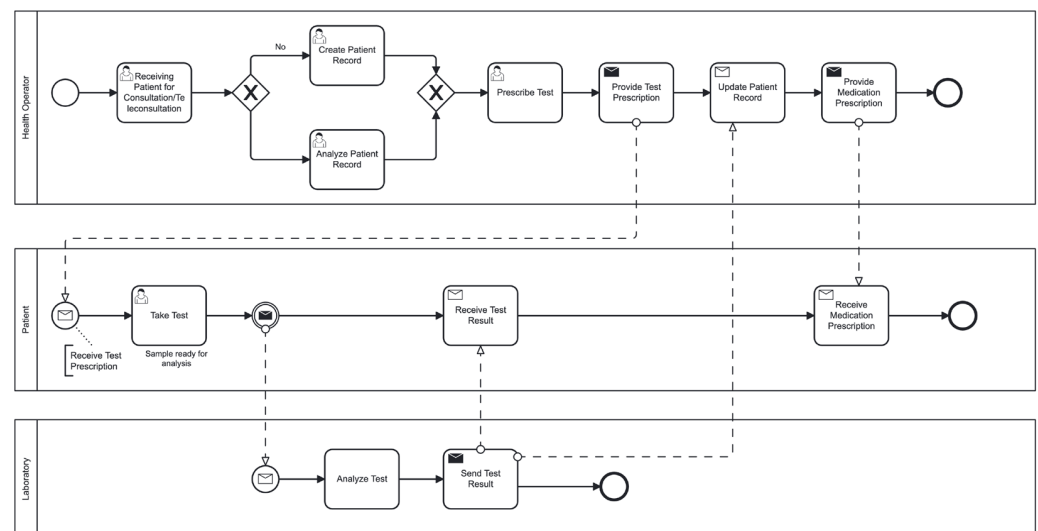


Figure 14. Pandemic Medication Planning Process Model.

Then, the health operator sends the form to the patient to keep it and depending on the symptoms he presents, he prescribes an RT-PCR (*Reverse Transcriptase Polymerase Chain Reaction*) test. The patient must take the RT-PCR test in a laboratory which must return the results to him within 48 h. The figure below shows the business process that we have just described.

Figure 15 presents *Dimensional Fact Model* of the *Pandemic Medication Planning* dataset. This diagram outlines the structure of the dataset, which captures the integration of task-level events with medical planning and prescription-related information. The dataset includes the following key dimensions (red-colored in Figure 15): (i) *Task_Name*; (ii) *Event_Type*; (iii) *Originator*; (iv) *Patient*; (v) *Test*; (vi) *Medication*; (vii) *Prescription*, with each record uniquely identified by an ID and timestamped via the *Timestamp* field. The selected measures are the count of distinct *Medication* and *Task_Name* values, respectively (blue-colored in Figure 15). The schema reflects the linkage between diagnostic procedures, patient-specific data, and therapeutic decisions, thereby enabling a coherent understanding of the medication planning process during pandemic response scenarios. By formalizing these relationships, the conceptual schema provides clarity on the dataset structure and enhances its applicability for evaluating clinical workflows, prescription compliance, and treatment decision-making dynamics in emergency health contexts.

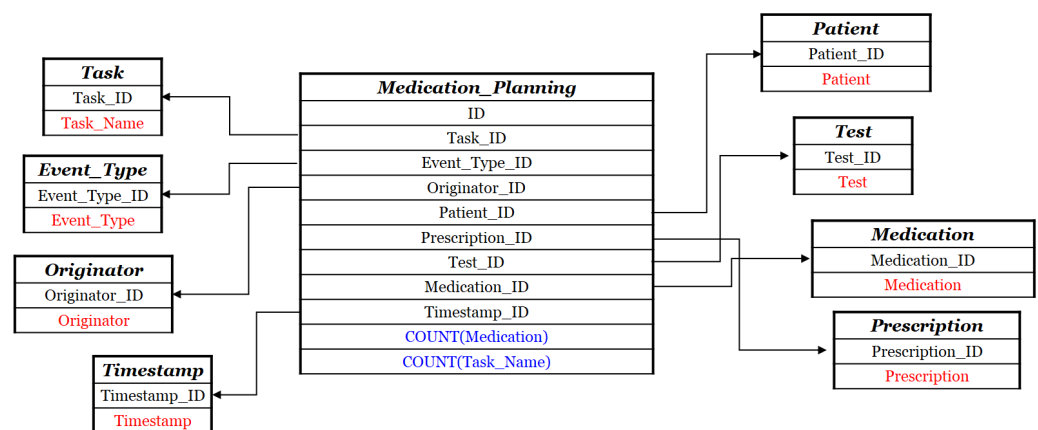


Figure 15. DFM over the Pandemic Medication Planning Dataset.

Finally, for medical pandemic fighting and prevention measures focusing on medical test and medication prescription management, the extracted dataset comprises key multidimensional attributes/measures such as *Task_Name*, specifying the nature of medical procedures undertaken; *Event_Type*, delineating the type of event recorded; *Originator*, identifying the initiator of each action; *Timestamp*, providing temporal context for events; *Patient*, identifying the individuals undergoing medical interventions; *Test*, detailing the diagnostic examinations conducted; *Medication*, specifying pharmaceutical interventions administered; and *Prescription*, documenting the prescribed treatments. Such a dataset enables rigorous analysis of medical intervention patterns during pandemics, facilitating insights into the effectiveness of preventive measures and the optimization of medical resource allocation. Figure 16 represents a sample of the *Pandemic Medication Planning* dataset.

Figure 17 shows an aggregation analysis bar plot representing the count of medications (attribute *Medication*) that have been prescribed to patients after taking each type of medical test (attribute *Test*).

Figure 18 shows a pivoting analysis by multi-attribute bar plot representing the count of medications (attribute *Medication*) that have been prescribed to patients after taking each type of medical test (attribute *Test*).

Task_Name	Event_Type	Originator	Timestamp	Patient	Test	Medication	Prescription
Prescribe medication	Completed	D8848646	3/1/2024 2:00:03 PM	P3153454	Molecular	Tocilizumab	PR0113213
Prescribe medication	Completed	D0128864	3/1/2024 2:00:03 PM	P3543135	Rapid Antigen	Tamiflu/Relenza	PR224544
Prescribe test	Dismissed	D0531548	18/1/2024 2:00:18 PM	P3153454	Screening	Tamiflu/Relenza	PR0113213
Prescribe test	Dismissed	D4843135	21/1/2024 2:00:21 PM	P5454641	Imaging	Antibiotic	PR0113213
Prescribe medication	Completed	D4843135	15/1/2024 2:00:15 PM	P3153454	Rapid Antigen	Steroid	PR224544
Prescribe medication	Dismissed	D3544651	6/1/2024 2:00:06 PM	P2313245	Serological	LMWH	PR0245433
Prescribe test	Dismissed	D4843135	30/1/2024 2:00:30 PM	P5454641	Pathogen Diagnostic	Steroid	PR244757
Prescribe medication	Completed	D4843135	31/1/2024 2:00:31 PM	P3153454	Blood	Tamiflu/Relenza	PR253443
Prescribe medication	Dismissed	D3544651	15/1/2024 2:00:15 PM	P3535644	Serological	Antibiotic	PR0113213
Prescribe test	Completed	D4843135	3/1/2024 2:00:03 PM	P2313245	Molecular	Tamiflu/Relenza	PR2154544
Prescribe test	Completed	D0654866	27/1/2024 2:00:27 PM	P5454641	Blood	Tocilizumab	PR0113213
Prescribe medication	Completed	D4843135	18/1/2024 2:00:18 PM	P3546846	Serological	Paracetamol/NSAIDs	PR0245433
Prescribe test	Completed	D0128864	6/1/2024 2:00:06 PM	P3545642	Imaging	Antibiotic	PR2154544
Prescribe test	Completed	D0531548	9/1/2024 2:00:09 PM	P3545642	Screening	Steroid	PR0245433
Prescribe medication	Dismissed	D0654866	1/1/2024 2:00:01 PM	P2313245	Pathogen Diagnostic	Antibiotic	PR253443
Prescribe medication	Dismissed	D0128864	27/1/2024 2:00:27 PM	P3545642	Serological	Tocilizumab	PR0113213
Prescribe test	Completed	D4843135	31/1/2024 2:00:31 PM	P3135431	Imaging	LMWH	PR0113213
Prescribe test	Completed	D3544651	18/1/2024 2:00:18 PM	P2313245	Molecular	Antibiotic	PR2154544

Figure 16. Sample from the Pandemic Medication Planning Dataset.

Count of Medication by Test

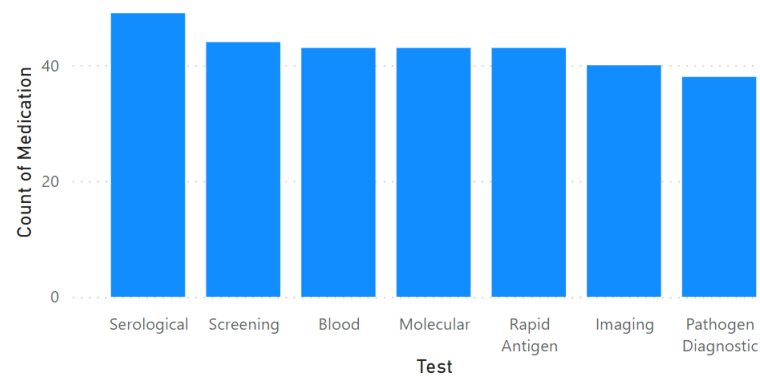


Figure 17. Aggregation Analysis of Medications Prescribed to Patients Based on Medical Tests.

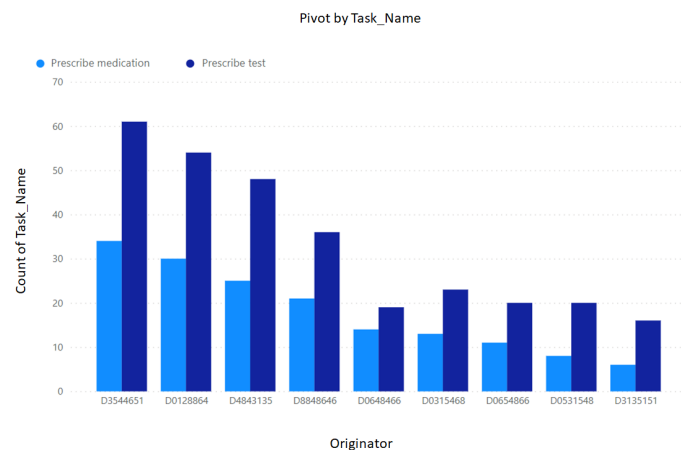


Figure 18. Pivoting Analysis of the Count of Medications Prescribed to Patients Based on Medical Tests.

Figure 19 shows the result of applying the clustering algorithm on the *Pandemic Medication Planning* dataset presented in Figure 16. The clustering is performed on the medications (attribute *Medication*) prescribed to patients (attribute *Patient*).

The results of this experimental campaign emphasize the adaptability of our framework in handling diverse pandemic-related data scenarios, including the integration of test outcomes. The analysis demonstrates that by applying multidimensional data exploration techniques, it is possible to extract meaningful insights into the relationships between testing data, resource allocation, and intervention strategies. When multiple tests are present

in the dataset, the framework effectively captures temporal trends, variations in testing frequency, and correlations with other pandemic-related activities. The aggregation and pivoting analysis allow decision-makers to track patterns in test results over time, assess the impact of testing policies, and identify clusters of similar behavioral responses. These findings highlight the robustness of our approach to managing dynamic and evolving pandemic datasets, supporting scalable and data-driven decision-making.

Medication and Patient (clusters)

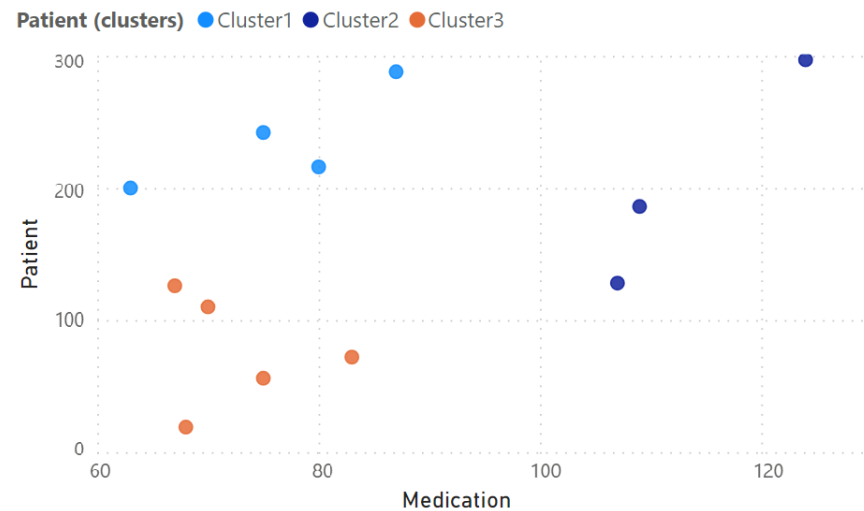


Figure 19. Clustering Results on the Pandemic Medication Planning Dataset.

5.4. Pandemic Case Surveillance

The business process model *Pandemic Case Surveillance* is depicted in Figure 20, which illustrates a tri-layered system comprising Case Reporting, Public Health Authority, and Case Notification. In the Case Reporting layer, entities such as hospitals, healthcare providers, and laboratories generate pandemic case reports in an integrated way and exchange pertinent data with local/state/territorial/tribal public health authorities. Subsequently, in the Public Health Authority layer, the latest pandemic case reports are exchanged with higher-level entities like the *Centers for Disease Control and Prevention* (CDC) and WHO. Finally, in the Case Notification layer, the received pandemic case reports are analyzed, and necessary pandemic measures are determined. These measures are then disseminated publicly for implementation, ensuring a coordinated response to the pandemic based on timely and accurate surveillance data.

Figure 21 presents *Dimensional Fact Model* of the *Pandemic Case Surveillance* dataset. This diagram outlines the structure of the dataset, which integrates critical event-level data related to case surveillance and public health measures during a pandemic. The dataset includes the following dimensions (red-colored in Figure 21): (i) *Task_Name*; (ii) *Event_Type*; (iii) *Originator*; (iv) *Timestamp*; (v) *Case*; (vi) *PublicHealthAuthority*; (vii) *HealthcareProvider*. Each record is uniquely identified by the ID and timestamped using the *Timestamp* field, ensuring precise tracking of events in the surveillance process. The selected measures are the count of distinct *Measure* and *Task_Name* values, respectively (blue-colored in Figure 21).

The schema highlights the connections between surveillance activities, case management, public health interventions, and healthcare provider involvement. This structure enables a comprehensive understanding of the pandemic response, tracking the identification and monitoring of cases, the implementation of health measures, and the roles of public health authorities and healthcare providers in managing the crisis. By formalizing these relationships, the conceptual schema facilitates the evaluation of case surveillance

processes, public health policy adherence, and healthcare provider coordination, thus improving preparedness and response during public health emergencies.

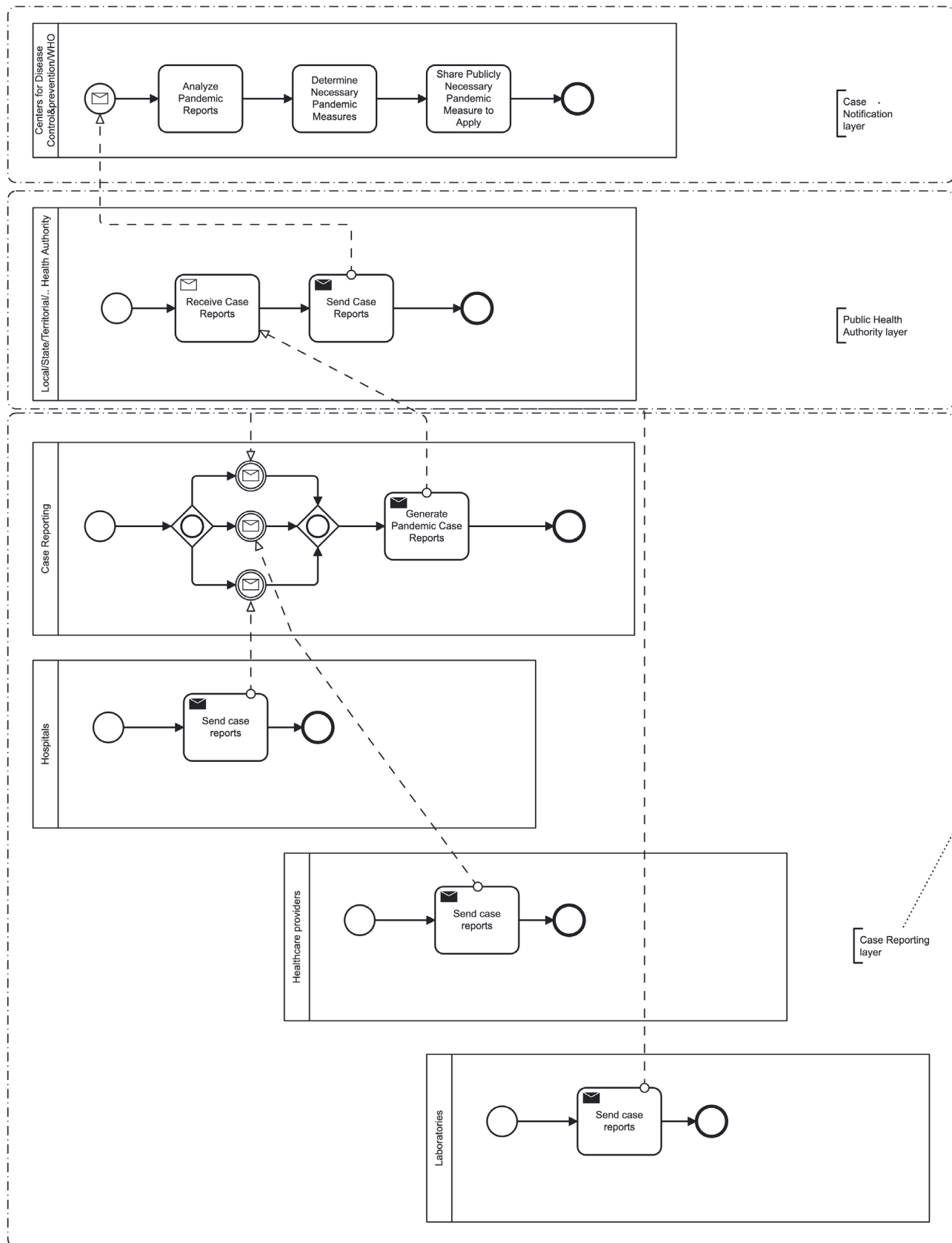


Figure 20. Pandemic Case Surveillance Process Model.

The *Pandemic Case Surveillance* dataset comprises critical attributes including: *Task_Name*, delineating the specific surveillance activities undertaken; *Event_Type*, categorizing the nature of recorded events; *Originator*, identifying the entity initiating each action; *Timestamp*, providing temporal context for events; *Case*, specifying the reported cases under surveillance; *Measure*, detailing the aggregation of a measured information value in the process of pandemic surveillance; *PublicHealthAuthority*, identifying the responsible health authority overseeing surveillance efforts; and *HealthcareProvider*, indicating the healthcare entities involved in case management. Such a dataset facilitates comprehensive analysis of pandemic surveillance strategies, enabling timely interventions and informed decision-making to mitigate the spread of infectious diseases. Figure 22 shows a sample of *Pandemic Case Surveillance* dataset described above.

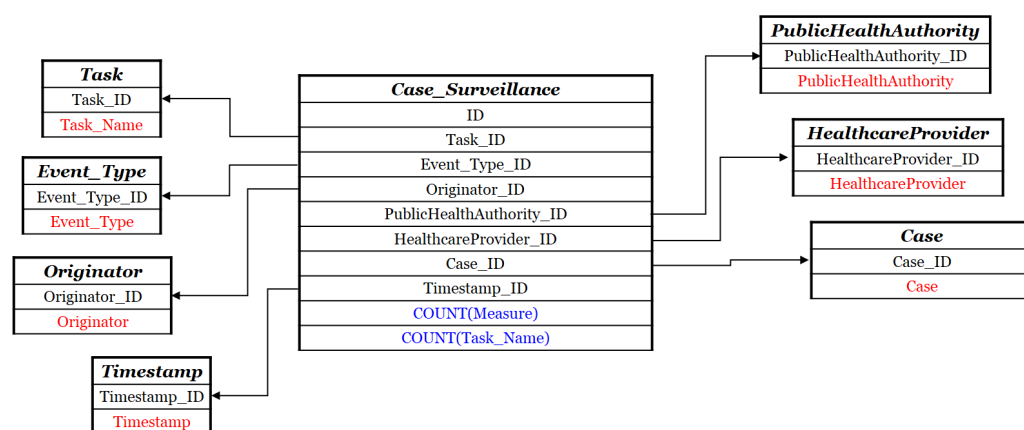


Figure 21. DFM over the Pandemic Case Surveillance Dataset.

Task_Name	Event_Type	Originator	Timestamp	Case	Measure	PublicHealthAuthority	HealthcareProvider
Generate case report	Completed	P0315468	21/1/2024 2:00:21 PM	West African Ebola Outbr	5455641	S3515516	P346416
Generate case report	Dismissed	P3544651	1/12/2024 2:00:12 PM	H1N1 Influenza Pandemi	3135151	L4164641	P311155
Authority send case report	Dismissed	P0531548	24/1/2024 2:00:12 PM	Zika Virus Epidemic	2456846	S0135445	P311155
Hospital Send case report	Dismissed	P3544651	15/1/2024 2:00:15 PM	West African Ebola Outbr	5455641	L3546464	P131565
Hospital Send case report	Completed	P0315468	21/1/2024 2:00:21 PM	West African Ebola Outbr	2810017	S5646844	P346416
Generate case report	Completed	P0128864	1/3/2024 2:00:03 PM	Zika Virus Epidemic	3215646	L3546464	P311155
Generate case report	Dismissed	P0128864	1/9/2024 2:00:09 PM	Middle East Respiratory S	2456846	T4863121	P135431
Authority send case report	Dismissed	P3135151	18/1/2024 2:00:18 PM	Ebola Outbreak	2810017	L0135435	P35413
Generate case report	Completed	P0654866	27/1/2024 2:00:27 PM	COVID-19 Pandemic	2452131	S0135445	P131565
Generate case report	Completed	P3135151	1/9/2024 2:00:09 PM	H1N1 Influenza Pandemi	2724125	S0135445	P315133
Generate case report	Dismissed	P0128864	1/6/2024 2:00:06 PM	Zika Virus Epidemic	2456846	T4646488	P31115
Hospital Send case report	Dismissed	P3544651	1/6/2024 2:00:06 PM	Zika Virus Epidemic	2810017	S5646844	P354135
Generate case report	Dismissed	P4843135	31/1/2024 2:00:31 PM	H1N1 Influenza Pandemi	3215646	S3515516	P311155
Authority send case report	Completed	P0531548	24/1/2024 2:00:12 PM	H1N1 Influenza Pandemi	2452131	L4164641	P31115
Hospital Send case report	Completed	P3544651	1/1/2024 2:00:01 PM	Zika Virus Epidemic	1315631	L7946563	P31115
Hospital Send case report	Dismissed	P3544651	1/12/2024 2:00:12 PM	Ebola Outbreak	2131351	L3546464	P135431
Authority send case report	Completed	P0128864	15/1/2024 2:00:15 PM	Middle East Respiratory S	5446461	T4646488	P135431
Generate case report	Completed	P0654866	1/6/2024 2:00:06 PM	Middle East Respiratory S	3546464	L3546464	P135431
Generate case report	Completed	P0531548	27/1/2024 2:00:27 PM	COVID-19 Pandemic	2810017	L0135436	P354135
Generate case report	Completed	P8848646	27/1/2024 2:00:27 PM	Middle East Respiratory S	2131351	S0135445	P35413
Generate case report	Completed	P0128864	1/3/2024 2:00:03 PM	Zika Virus Epidemic	2131351	S0135445	P311155

Figure 22. Sample from the Pandemic Case Surveillance Dataset.

Figure 23 shows a bar plot representing the count of the aggregate measure value (attribute *Measure*) related to each pandemic case (attribute *Case*).

In Figure 24, we show a stacked column bar plot of the count of tasks (attribute *Task_Name*) executed by each public health authority while the process of pandemic surveillance based on the different pandemic cases (attribute *Case*).

Figure 25 shows the result of applying the clustering algorithm on the *Pandemic Case Surveillance* dataset presented in Figure 22. The clustering is performed on the previous known pandemic cases handled by public healthcare authorities (attribute *Case*) by the aggregation value of the measure (attribute *Measure*) healthcare providers.

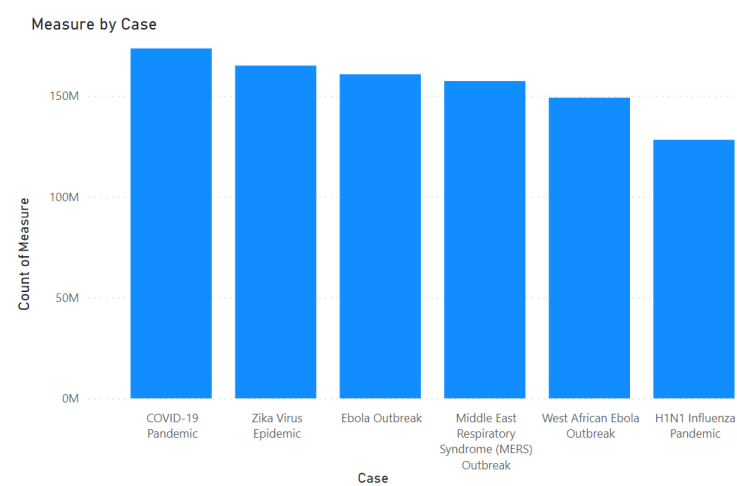


Figure 23. Aggregation Analysis of Measure Value by Pandemic Case.

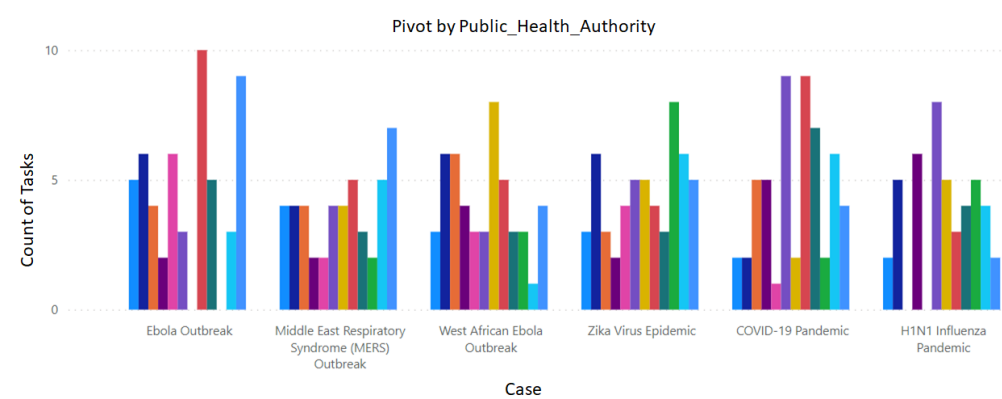


Figure 24. Pivoting Analysis of Tasks Executed by each Public Health Authority based on the different Pandemic Cases.

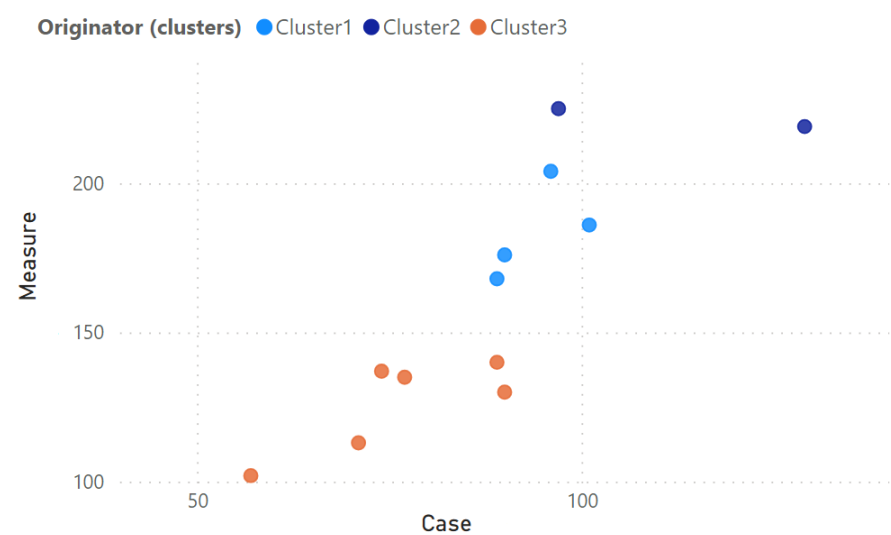


Figure 25. Clustering Results on the Pandemic Case Surveillance Dataset.

5.5. Pandemic Vaccination Planning

The business process model *Pandemic Vaccination Planning* is shown in Figure 26, which outlines a structured approach based on age, medical condition, and community impact. Individuals aged over 65 automatically qualify for vaccination. For those aged between 50 and 65 with a medical condition or residing in a disproportionately affected community, vaccination is also recommended. Further stratification involves individuals

aged 18–49 without insurance or underinsured, residing in affected communities, and with medical conditions, who are prioritized for vaccination. However, due to limited vaccine availability, individuals not meeting these criteria or falling outside the specified age brackets are placed on a waitlist. This model ensures a systematic allocation of vaccines, prioritizing vulnerable demographics while managing resource constraints effectively.

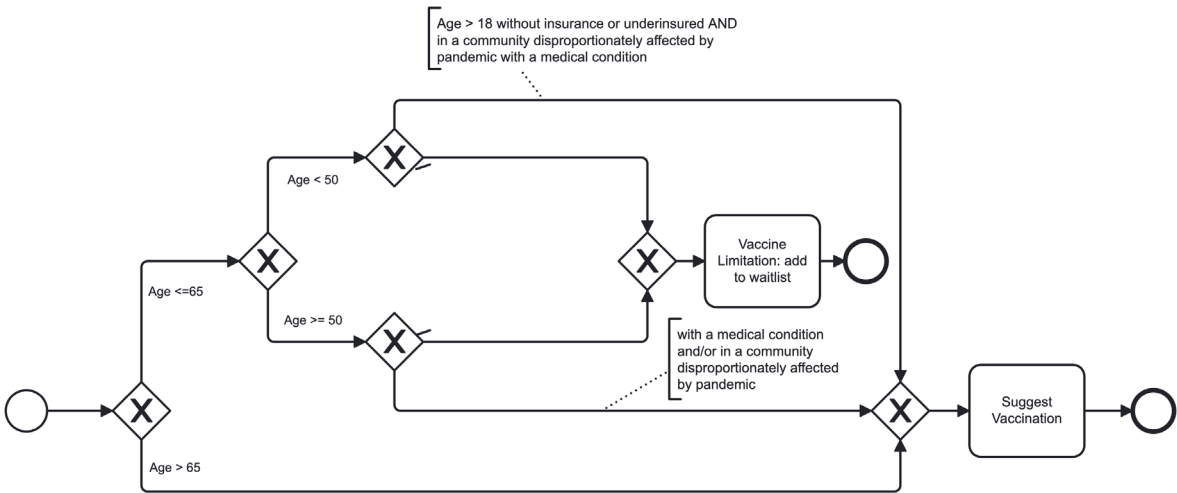


Figure 26. Pandemic Vaccination Planning Process Model.

Figure 27 presents *Dimensional Fact Model* of the *Pandemic Vaccination Planning* dataset. This diagram depicts the structure of the dataset, which captures task-level events related to vaccine administration planning, prioritization, and allocation during a pandemic scenario. The model includes the following core dimensions (red-colored in Figure 27): (i) *Task_Name*; (ii) *Event_Type*; (iii) *Originator*; (iv) *Timestamp*; (v) *Age*; (vi) *MedicalCondition*; (vii) *Vaccine*; (viii) *Waitlist*. The selected measures are the count of distinct *Originator* value (blue-colored in Figure 27). Each record is uniquely identified by the ID and timestamped through the *Timestamp* field, allowing for chronological tracking of vaccination-related events.

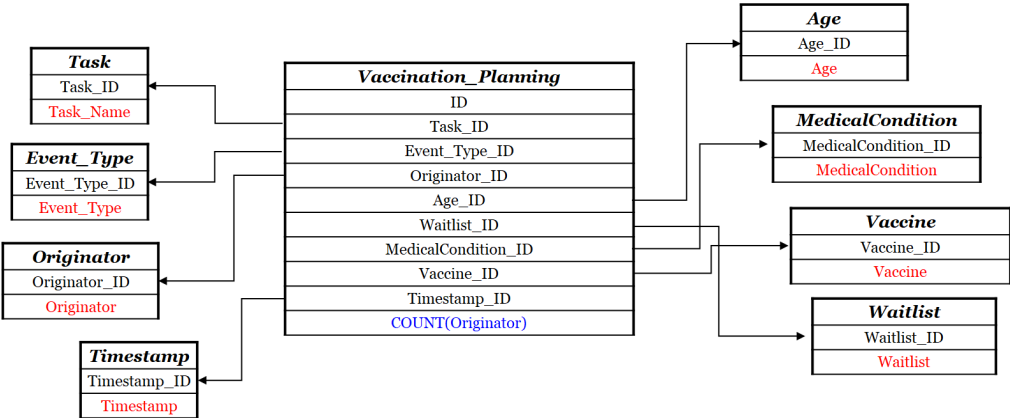


Figure 27. DFM over the Pandemic Vaccination Planning Dataset.

The schema reflects the integration of demographic and clinical criteria, such as *Age* and *MedicalCondition* with logistical aspects of vaccination, including *Vaccine* and *Waitlist* status. This design enables detailed analysis of prioritization strategies, vaccine distribution logistics, and eligibility determination. By establishing clear relationships among these factors, the conceptual schema supports data-driven decision-making in vaccine rollout strategies, facilitates equitable access assessments, and enhances understanding of vaccination planning dynamics during public health crises.

In the case of pandemic, vaccination planning constitutes a critical resource for informing vaccination strategies and prioritization efforts. It encompasses essential attributes such as *Task_Name*, specifying the particular vaccination planning activities undertaken; *Event_Type*, categorizing the nature of recorded events; *Originator*, identifying the initiator of each action within the vaccination planning process; *Timestamp*, providing temporal context for events; *Age*, indicating the age demographic of individuals within the planning framework; *MedicalCondition*, detailing any underlying health conditions relevant to vaccination prioritization; *Vaccine*, denoting individuals eligible for vaccination; and *Waitlist*, identifying individuals who are recommended to be placed on a vaccination waitlist pending availability. Such a dataset enables systematic analysis and optimization of vaccination distribution strategies, ensuring efficient and equitable allocation of limited vaccine resources during pandemics. In Figure 28, we provide a sample of the *Pandemic Vaccination Planning* dataset.

Task_Name	Event_Type	Originator	Timestamp	Age	MedicalCondition	Vaccine	Waitlist
Take vaccine	Dismissed	P3135151	30/1/2024 2:00:30 PM	<50	No	No	No
Not take vaccine	Dismissed	P4843135	1/9/2024 2:00:09 PM	<50	No	Yes	Yes
Take vaccine	Dismissed	P8848646	1/6/2024 2:00:06 PM	<50	No	Yes	Yes
Not take vaccine	Dismissed	P3544651	15/1/2024 2:00:15 PM	<50	Yes	No	Yes
Take vaccine	Dismissed	P3544651	27/1/2024 2:00:27 PM	<50	Yes	No	No
Take vaccine	Dismissed	P0531548	30/1/2024 2:00:30 PM	<50	No	Yes	No
Not take vaccine	Completed	P3135151	21/1/2024 2:00:21 PM	<50	Yes	Yes	Yes
Take vaccine	Completed	P8848646	15/1/2024 2:00:15 PM	<50	Yes	No	No
Take vaccine	Completed	P4843135	27/1/2024 2:00:27 PM	<50	Yes	No	No
Take vaccine	Completed	P8848646	27/1/2024 2:00:27 PM	<50	No	Yes	Yes
Take vaccine	Dismissed	P0654866	18/1/2024 2:00:18 PM	<50	Yes	Yes	Yes
Take vaccine	Completed	P4843135	1/9/2024 2:00:09 PM	<50	Yes	Yes	Yes
Take vaccine	Dismissed	P0128864	24/1/2024 2:00:12 PM	<50	No	No	No
Take vaccine	Completed	P0128864	1/6/2024 2:00:06 PM	<50	Yes	Yes	Yes
Not take vaccine	Completed	P0315468	15/1/2024 2:00:15 PM	<50	No	No	Yes
Take vaccine	Completed	P3544651	15/1/2024 2:00:15 PM	<50	No	Yes	Yes
Take vaccine	Completed	P4843135	18/1/2024 2:00:18 PM	<50	No	Yes	No

Figure 28. Sample from the Pandemic Vaccination Planning Dataset.

Figure 29 shows a bar plot representing the count of patients (attribute *Originator*) eligible for vaccination who have performed the vaccination and of those who did not (attribute *Task_Name*).

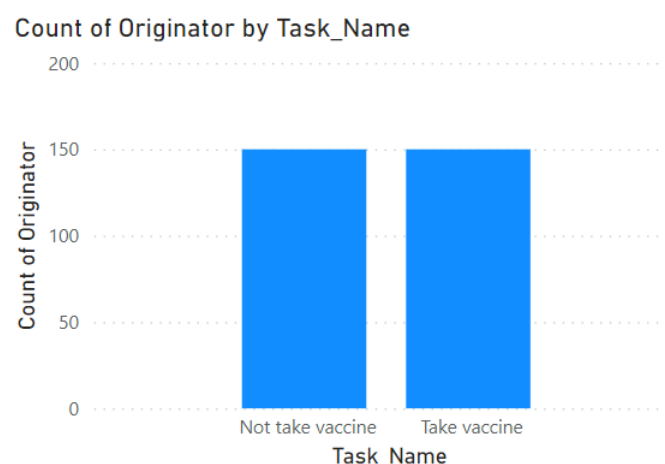


Figure 29. Aggregation Analysis of Patients Eligible for Vaccination that Vaccinated/Did Not.

In Figure 30, we show a stacked column bar plot of the count of patients (attribute *Originator*) eligible for vaccination (attribute *Vaccination*) based on their age (attribute *Age*).

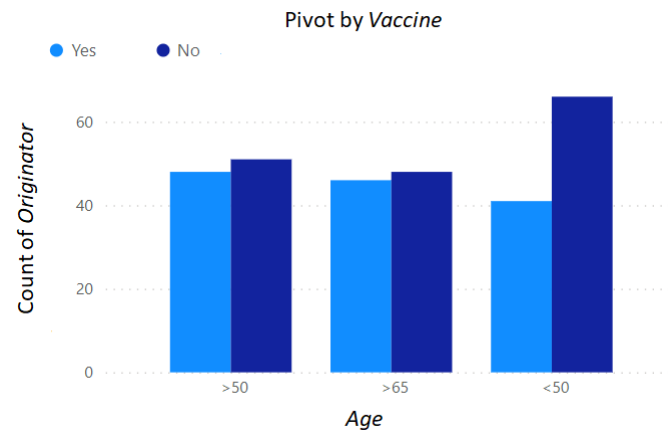


Figure 30. Pivoting Analysis of Patients Eligible for Vaccination based on their Age.

Figure 31 shows the result of applying the clustering algorithm on *Pandemic Vaccination Planning* dataset presented in Figure 28. The clustering is performed on the age of patients (attribute *Age*) eligible for vaccination (attribute *Vaccine*).

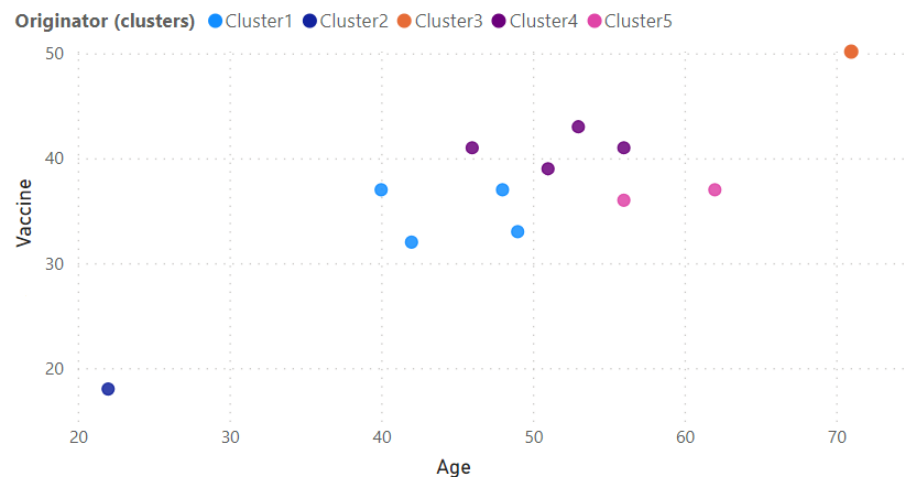


Figure 31. Clustering Results on the Pandemic Vaccination Planning Dataset.

For the purpose of *behavioral pattern analysis*, the PROTECTION framework integrates advanced data aggregation, multidimensional pivoting, and adaptive clustering methodologies to facilitate the scalable and interpretable extraction of behavioral patterns, temporal trends, and critical operational insights from complex, high-dimensional pandemic-related datasets. While these techniques are commonly found in conventional business intelligence (BI) platforms, PROTECTION re-engineers them within a context-aware, dynamic analytical architecture designed to support real-time data streams, semantic filtering, and adaptive analytical workflows.

Technically, the framework leverages a streaming pivot engine, implemented using *Apache Flink*, which supports *low-latency transformation of high-frequency input streams into structured summary views*. A *semantic annotation module* enhances data with domain-specific ontologies (e.g., *HL7 FHIR*, *COVID-19 OWL*), enabling guided exploration and context-driven feature selection. The clustering component employs an incremental density-based algorithm (i.e., *adaptive DBSCAN*) capable of recalibrating cluster boundaries on-the-fly in response to temporal and spatial changes in data distributions, without requiring full re-computation.

Unlike traditional BI systems such as *Pentaho* or *Tableau*, which rely on static pivot tables and pre-defined clustering parameters, PROTECTION introduces runtime adaptation mechanisms that allow analytical components to evolve in response to real-time data

characteristics. This includes the use of semantic constraints to refine data slicing and the dynamic adjustment of clustering thresholds based on concept drift and temporal density shifts.

The efficacy of these techniques was evaluated through a series of controlled experiments on two real-world datasets:

- COVID-19 Mobility and Contact Tracing Dataset [71]
- Syndromic Surveillance and Social Media Signal Datasets [72,73]

Each dataset was streamed over a 30-day simulation to mimic real-time ingestion. Metrics used to assess performance included detection latency, clustering accuracy, and semantic interpretability.

On the COVID-19 Mobility and Contact Tracing Dataset, PROTECTION adaptive clustering module identified deviations in population mobility patterns correlated with evolving public health policies with an average detection latency of 4.3 h, compared to 6.6 h using a *static DBSCAN baseline* (35% improvement, $p < 0.01$).

On the other hand, for the Syndromic Surveillance and Social Media Signal Dataset, semantic-guided pivoting and annotation-based feature selection improved trend classification *F1-scores* (e.g., public sentiment on mask-wearing) from 0.72 (*static BI approach*) to 0.86, while domain expert evaluation rated the interpretability of insights at 4.6/5, versus 3.8/5 for the baseline system.

These results demonstrate that PROTECTION unified approach substantially enhances the relevance, timeliness, and contextual fidelity of extracted insights in dynamic, high-frequency operational environments. Moreover, it is important to clarify that in PROTECTION, the quantitative validation results obtained through analytics are *not automatically fed back into the BPMN process models* for structural refinement. Instead, these results serve as actionable inputs for *healthcare decision-makers*. In this way, analytics outcomes complement the BPMN process representations by enriching *decision-making* with *data-driven insights*, rather than dynamically modifying BPMN themselves.

Nevertheless, we acknowledge that process mining reaches its full potential only when applied to authentic event logs that may reveal unexpected, unconventional, or non-compliant behaviors that are not anticipated in preconceived models. While the synthetic case studies have been essential to ensure reproducibility and to address privacy restrictions, future research direction will focus on deploying the PROTECTION framework on anonymized event logs collected in collaboration with healthcare institutions. While alternative methodologies, such as probabilistic models, Bayesian inference, and uncertain data analytics [74–76], offer additional depth for managing incomplete or noisy data, their integration into the current PROTECTION framework remains a direction for future work. This will allow us to capture unforeseen deviations in real pandemic workflows, thereby extending the validation of KDDI process modeling beyond controlled datasets and further strengthening the practical impact of our proposal.

6. Focused Vertical Use Case: ICU Resource Planning During a COVID-19 Wave Surge

In this Section, we provide a detailed vertical scenario representative of the platform. We have incorporated a focused use case in the revised manuscript: ICU resource planning during a COVID-19 wave surge. This scenario exemplifies the PROTECTION framework ability to model, manage, and mine data-centric pandemic control processes by integrating multiple domains through executable BPMN workflows.

In this use case, a regional health authority leverages the PROTECTION platform to dynamically assess and allocate *Intensive Care Unit* (ICU) resources in response to an escalating COVID-19 outbreak. The scenario involves a coordinated, multi-actor process

spanning five key pandemic control domains, each contributing specific data entities, event types, and decision logic to the composite process model:

Behavioral Infection Control: Population mobility data and adherence metrics to masking and isolation guidelines are continuously ingested and visualized. These behavioral indicators, modeled as BPMN events and exclusive gateways, influence predicted transmission rates, triggering escalation paths in the process when thresholds are breached.

Environmental Pandemic Control: Environmental monitoring nodes report real-time air quality and viral load indices from high-risk indoor settings (e.g., *hospitals, transport hubs*). These inputs are modeled as intermediate message events that influence risk stratification and determine hotspot classification within the process logic.

Pandemic Case Surveillance: Confirmed case reports, contact tracing data, and epidemiological indicators such as R-values are ingested via structured message flows from public health authorities. These data directly influence the triggering of ICU surge protocols, represented by BPMN event sub-processes conditioned on case velocity and geographic clustering.

Pandemic Medication Planning: Medication inventory levels (e.g., *sedatives, antivirals*) and prescription patterns are monitored to ensure therapeutic support for ICU patients. Shortages automatically trigger compensation tasks (e.g., *alternative drug sourcing*), modeled through data-driven inclusive gateways.

Pandemic Vaccination Planning: Vaccination coverage among high-risk and ICU-eligible populations is incorporated to forecast protection levels. Low coverage prompts prioritization sub-processes to expedite booster campaigns, directly modeled as conditional event sub-processes within the BPMN structure.

To enable continuous monitoring and refinement of ICU resource planning during a COVID-19 wave surge, the PROTECTION platform integrates core process mining techniques, namely process discovery, conformance checking, and process enhancement, into its execution and analysis framework. These methods support adaptive decision-making and operational transparency across the five interconnected pandemic control domains.

1. *Process Discovery:* Process discovery aims to extract executable process models from real-time event data collected during the enactment of pandemic-related workflows. In the ICU resource planning context, heterogeneous event logs from surveillance systems, hospital resource tracking, medication inventories, and public health portals are consolidated into a unified event stream. Technically, the platform employs algorithms such as the α -algorithm, Heuristic Miner, and Inductive Miner to reconstruct as-is BPMN models from logged execution traces. For instance:
 - In Pandemic Case Surveillance, event logs of confirmed cases and contact tracing records are used to derive empirical transmission chains and escalation workflows.
 - In Medication Planning, prescription fulfillment and drug stock replenishment logs are mined to uncover bottlenecks and implicit resource allocation practices not captured in the original model.
 - These discovered models provide a ground-truth representation of actual operational behavior, enabling the comparison against intended process specifications.
2. *Conformance Checking:* Conformance checking evaluates the alignment between the discovered (or actual) process behavior and the reference process model encoded in the BPMN workflows. This is particularly critical in healthcare, where non-compliant behavior can have direct consequences on patient outcomes. Within the PROTECTION platform, conformance checking is performed using:
 - Token-based replay to quantify deviations (e.g., *missing ICU bed allocation steps, unauthorized medication usage*).

- Alignment-based techniques, which calculate fitness, precision, and generalization metrics between event logs and model pathways.

For example:

- In the Vaccination Planning domain, if the event log indicates delays in initiating booster campaigns in high-risk regions, the system flags this as a temporal deviation from the expected conditional sub-process.
- In Behavioral Infection Control, discrepancies between observed mobility data and mandated isolation protocols are detected as violations of the process model, triggering compliance alerts.

Such conformance analyses are used not only for auditing and regulatory reporting but also to generate actionable alerts for operational correction.

3. **Process Enhancement:** Process enhancement uses observed event data to improve existing process models by integrating new behavior, optimizing performance, or annotating models with predictive metrics.

In the ICU planning scenario, enhancement is applied in several ways:

- Performance annotation augments BPMN models with time-based statistics such as average ICU occupancy transition times, drug sourcing delays, or case escalation durations.
- Predictive modeling is integrated using process-aware machine learning (e.g., decision trees, LSTMs) to forecast ICU bed shortages based on recent case velocity and environmental indicators.
- Resource enhancement updates gateway logic or sub-process parameters based on the inferred dependencies, for example, adjusting medication replenishment thresholds or modifying ICU admission criteria in response to changing case clusters.

This dynamic refinement ensures the process model remains aligned with evolving pandemic dynamics, supporting scenario simulations and stress testing of healthcare infrastructure.

Through this use case, the PROTECTION platform demonstrates how complex, high-stakes healthcare processes can be modeled using BPMN enriched with domain-specific, real-time data. The system enables hospital administrators and public health officials to pose complex analytical queries (e.g., “Which regions will face ICU bed shortages within the next 10 days given current case trajectories and medication availability?”) and receive process-aware visualizations and alerts.

This scenario, built upon real-world parameters and simulated surge data via a BPMN model and dataset that not only illustrates the utility of the platform in a concrete domain-specific application but also provides an extensible template for similar use cases in pandemic preparedness and response planning.

The process model shown in Figure 32 illustrates the integration of five pandemic control domains: Behavioral Infection Control, Environmental Pandemic Control, Pandemic Case Surveillance, Pandemic Medication Planning, and Pandemic Vaccination Planning within the PROTECTION framework. Real-time data from behavioral compliance, environmental viral indicators, surveillance systems, medication inventories, and vaccination coverage dynamically influence the execution paths of the process. The BPMN diagram includes conditional events, data-driven gateways, and escalation sub-processes to model ICU surge response planning. This figure demonstrates the platform capability to operationalize complex, multi-domain health processes and support data-centric decision-making under pandemic pressure.

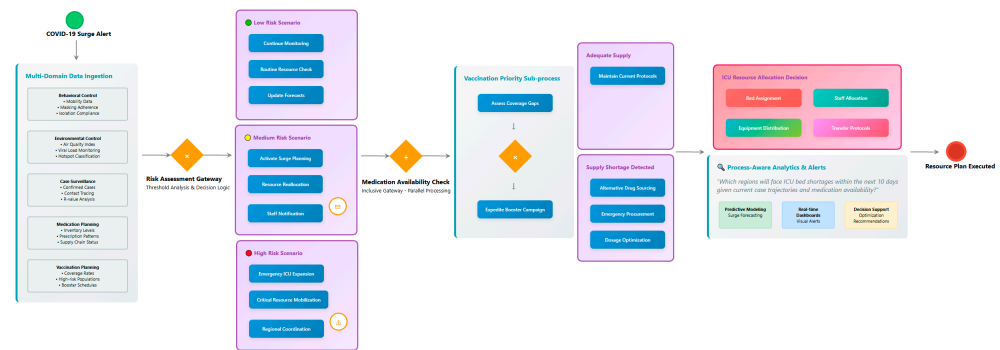


Figure 32. BPMN-based vertical use case for ICU Resource Planning during a COVID-19 Wave Surge.

7. Conclusions and Future Work

This paper has proposed PROTECTION, an innovative data-centric process-modeling-managing-and-mining framework for pandemic control and prevention, grounded in the proposed KDDI processes paradigm. The COVID-19 pandemic outbreak highlighted the limitations of current healthcare information systems. Addressing the gap exposed by the COVID-19 experience, PROTECTION aims to enhance the resilience and adaptability of healthcare systems in the face of such crises.

In particular, PROTECTION leverages BPMN as a formal method for supporting the structured representation of healthcare and clinical workflows. BPMN enables healthcare organizations to model and visualize complex KDDI processes in a standardized and interoperable format, which represents a critical advantage when dealing with rapidly evolving pandemic scenarios. In fact, the use of BPMN not only supports clarity in design but also facilitates real-time process tracking and adaptive modification, capabilities that are central to PROTECTION design. The framework functionalities, as demonstrated through several case studies, illustrate its potential to optimize pandemic control and prevention efforts. By interpreting real-world-based clinical data and integrating it with BPMN-modeled processes, PROTECTION supports enhanced decision-making and operational efficiency.

Moving forward, future work will focus on further experimentally testing and refining the capabilities of PROTECTION, expanding its application across different healthcare contexts, and improving its real-time adaptability to evolving pandemic scenarios. The ongoing development and validation of this framework are expected to contribute significantly to the creation of more agile, data-driven healthcare systems capable of mitigating the impacts of future global health crises. On the other hand, *privacy and security* (e.g., [77,78]) and *uncertain data management and analytics* (e.g., [79,80]) are relevant research aspects that deserve further attention during next years.

Author Contributions: Conceptualization, A.C., C.C., E.F. and P.T.; methodology, A.C., C.C., E.F. and P.T.; software, I.B.; validation, A.C. and I.B.; formal analysis, A.C. and C.C.; investigation, A.C. and C.C.; resources, C.C.; data curation, A.C. and I.B.; writing—original draft preparation, A.C., I.B., C.C., E.F. and P.T.; writing—review and editing, A.C. and C.C.; visualization, A.C. and I.B.; supervision, A.C. and C.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Datasets available on request from the authors.

Acknowledgments: This research is supported by the ICSC National Research Centre for High Performance Computing, Big Data and Quantum Computing within the NextGenerationEU program (Project Code: PNRR CN00000013).

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Beheshti, A.; Schiliro, F.; Ghodrathnama, S.; Amouzgar, F.; Benatallah, B.; Yang, J.; Sheng, Q.Z.; Casati, F.; Motahari-Nezhad, H.R. iProcess: Enabling IoT Platforms in Data-Driven Knowledge-Intensive Processes. In Proceedings of the Business Process Management Forum, BPM (Forum) 2018, Sydney, Australia, 9–14 September 2018; pp. 108–126.
- Apté, C.; Hong, S.J.; Natarajan, R.; Pednault, E.P.D.; Tipu, F.; Weiss, S.M. Data-Intensive Analytics for Predictive Modeling. *IBM J. Res. Dev.* **2003**, *47*, 17–23. [\[CrossRef\]](#)
- Savas Ilgi, G.; Etikan, I.; Ever, Y.K. Developing a Machine Learning Algorithm to Determine COVID-19 Contamination in Different Age Groups and Comparing Statistical Algorithms and Learning Data. *IEEE Access* **2024**, *12*, 117461–117470. [\[CrossRef\]](#)
- Khan, M.A.; Latif, K.F.; Shahid, S.; Shah, S.A. Understanding Knowledge Leadership in Improving Team Outcomes in the Health Sector: A COVID-19 Study. *Bus. Process Manag. J.* **2024**, *30*, 63–83. [\[CrossRef\]](#)
- Gonçalves, J.C.A.R.; Baião, F.A.; Santoro, F.M.; Guizzardi, G. A Cognitive BPM Theory for Knowledge-Intensive Processes. *Bus. Process Manag. J.* **2023**, *29*, 465–488. [\[CrossRef\]](#)
- Nudurupati, S.S.; Tebboune, S.; Garengo, P.; Daley, R.; Hardman, J. Performance Measurement in Data Intensive Organisations: Resources and Capabilities for Decision-Making Process. *Prod. Plan. Control* **2024**, *35*, 373–393. [\[CrossRef\]](#)
- Estrada-Torres, B.; Richetti, P.H.P.; Del-Rio-Ortega, A.; Baiao, F.A.; Resinas, M.; Santoro, F.M.; Ruiz-Cortés, A. Measuring Performance in Knowledge-Intensive Processes. *ACM Trans. Internet Technol.* **2019**, *19*, 1–26. [\[CrossRef\]](#)
- Seidel, A.; Haarmann, S. Decision Support for Knowledge-Intensive Processes. In Proceedings of the 14th Central European Workshop on Services and their Composition, ZEUS 2022, Bamberg, Germany, 24–25 February 2022; pp. 20–29.
- Niu, B.; Wang, L.; Yu, X.; Feng, B. Data-Driven Analysis of Digital Entrepreneurship in Medical Supply Resilience Confronting the COVID-19 Epidemic. *Inf. Process. Manag.* **2024**, *61*, 103502. [\[CrossRef\]](#)
- Saleh, S.N. Enhancing Multilabel Classification for Unbalanced COVID-19 Vaccination Hesitancy Tweets Using Ensemble Learning. *Comput. Biol. Med.* **2025**, *184*, 109437. [\[CrossRef\]](#)
- Kumar, S.; Garg, S.; Muhuri, P.K. A Stratified Review of COVID-19 Infection Forecasting and an Efficient Methodology using Multiple Domain-based Transfer Learning. *Expert Syst. Appl.* **2025**, *262*, 125277. [\[CrossRef\]](#)
- Kir, H.; Erdogan, N. A Knowledge-Intensive Adaptive Business Process Management Framework. *Inf. Syst.* **2021**, *95*, 101639. [\[CrossRef\]](#)
- Czvetkó, T.; Kummer, A.; Ruppert, T.; Abonyi, J. Data-Driven Business Process Management-Based Development of Industry 4.0 Solutions. *CIRP J. Manuf. Sci. Technol.* **2022**, *36*, 117–132. [\[CrossRef\]](#)
- Chakraborty, C.; Abougren, A.N. Intelligent Internet of Things and Advanced Machine Learning Techniques for COVID-19. *EAI Endorsed Trans. Pervasive Health Technol.* **2021**, *7*, 1–14. [\[CrossRef\]](#)
- Vanani, I.R.; Taghavifard, M.T.; Yalpanian, M.A. Predictive Analytics Solution for Digital Capabilities Identification Towards Business Performance Improvement. *SN Comput. Sci.* **2025**, *6*, 85. [\[CrossRef\]](#)
- Boonsoonthosattit, G.; Vongbunpong, S.; Chonsawat, N.; Chanpuypetch, W. Development of a Hybrid AHP-TOPSIS Decision-Making Framework for Technology Selection in Hospital Medication Dispensing Processes. *IEEE Access* **2024**, *12*, 2500–2516. [\[CrossRef\]](#)
- Wieckowski, J.; Salabun, W. Supporting Multi-Criteria Decision-Making Processes with Unknown Criteria Weights. *Eng. Appl. Artif. Intell.* **2025**, *140*, 109699. [\[CrossRef\]](#)
- Muhunzi, D.; Kitambala, L.; Mashauri, H.L. Big Data Analytics in the Healthcare Sector: Opportunities and Challenges in Developing Countries. A Literature Review. *Health Inform. J.* **2024**, *30*, 14604582241294217. [\[CrossRef\]](#)
- Nauman, M.; Almadhor, A.S.; Albekairi, M.; Ansari, A.R.; Fayyaz, M.A.B.; Nawaz, R. The Role of Big Data Analytics in Revolutionizing Diabetes Management and Healthcare Decision-Making. *IEEE Access* **2025**, *13*, 10767–10785. [\[CrossRef\]](#)
- National Institutes of Health. Open-Access Data and Computational Resources to Address COVID-19. Available online: <https://datascience.nih.gov/covid-19-open-access-resources> (accessed on 31 December 2024).
- The World Health Organization. Global Research on Coronavirus Disease (COVID-19). Available online: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov> (accessed on 31 December 2024).
- Istituto Superiore di Sanità. Epidemiology for Public Health. Available online: <https://www.epicentro.iss.it/coronavirus/> (accessed on 31 December 2024).
- European Centre for Disease Prevention and Control. Coronavirus Threats and Outbreaks: COVID-19 Pandemic. Available online: <https://www.ecdc.europa.eu/en/covid-19-pandemic> (accessed on 31 December 2024).
- Johns Hopkins University. COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE). Available online: <https://github.com/CSSEGISandData/COVID-19> (accessed on 31 December 2024).
- OpenICPSR. The COVID-19 Data Repository. Available online: <https://www.openicpsr.org/openicpsr/covid19> (accessed on 31 December 2024).
- Carbon Health and Braid Health. Coronavirus Disease 2019 Clinical Data Repository. Available online: <https://covidclinicaldata.org> (accessed on 31 December 2024).

27. Kaggle. Data Science for COVID-19 (DS4C) in South Korea. Available online: <https://www.kaggle.com/kimjihoo/coronavirusdataset> (accessed on 31 December 2024).
28. HHS Technology Group. COVID-19 Research Database. Available online: <http://hhstechgroup.com/covid-19-research-database-partners-with-the-hhs-technology-group-scale-an-open-research-database/> (accessed on 31 December 2024).
29. National Institutes of Health. Coronavirus Disease 2019 (COVID-19) Treatment Guidelines. Available online: <https://pubmed.ncbi.nlm.nih.gov/34003615/> (accessed on 31 December 2024).
30. World Health Organization. Country & Technical Guidance—Coronavirus Disease (COVID-19). Available online: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance-publications> (accessed on 31 December 2024).
31. European Respiratory Society. COVID-19: Guidelines and Recommendations Directory. Available online: <https://www.ersnet.org/covid-19/covid-19-guidelines-and-recommendations-directory/> (accessed on 31 December 2024).
32. Campbell, H.; Hotchkiss, R.; Bradshaw, N.; Porteous, M. Integrated Care Pathways. *Br. Med. J.* **1998**, *316*, 133–137. [CrossRef]
33. Combi, C.; Keravnou-Papailiou, E.; Shahar, Y. *Temporal Information Systems in Medicine*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2010.
34. Peleg, M. Computer-Interpretable Clinical Guidelines: A Methodological Review. *J. Biomed. Inform.* **2013**, *46*, 744–763. [CrossRef]
35. Wright, A.; Sittig, D.F.; Ash, J.S.; Feblowitz, J.; Meltzer, S.; McMullen, C.; Guappone, K.P.; Carpenter, J.; Richardson, J.E.; Simonaitis, L.; et al. Development and Evaluation of a Comprehensive Clinical Decision Support Taxonomy: Comparison of Front-End Tools in Commercial and Internally Developed Electronic Health Record Systems. *J. Am. Med. Inform. Assoc.* **2011**, *18*, 232–242. [CrossRef]
36. Peleg, M. Guidelines and Workflow Models. In *Clinical Decision Support: The Road Ahead*; Elsevier: San Diego, CA, USA, 2007; pp. 281–306.
37. Quaglini, S.; Stefanelli, M.; Lanzola, G.; Caporusso, V.; Panzarasa, S. Flexible Guideline-Based Patient Careflow Systems. *Artif. Intell. Med.* **2001**, *22*, 65–80. [CrossRef]
38. Rentes, V.C.; de Pádua, S.I.D.; Coelho, E.B.; de Camargo Teixeira Cintra, M.A.; Ilana, G.G.F.; Rozenfeld, H. Implementation of a Strategic Planning Process Oriented towards Promoting Business Process Management (BPM) at a Clinical Research Centre (CRC). *Bus. Process Manag. J.* **2019**, *25*, 707–737. [CrossRef]
39. Lario, R.F.; Soley, R.; White, S.; Butler, J.; Del Fiol, G.; Eilbeck, K.; Huff, S.M.; Kawamoto, K. The Business Process Management for Healthcare (BPM + Health) Consortium: Motivation, Methodology, and Deliverables for Enabling Clinical Knowledge Interoperability (CKI). *J. Am. Med. Inform. Assoc.* **2024**, *31*, 797–808. [CrossRef] [PubMed]
40. Chesani, F.; Mello, P.; Montali, M. Abductive Reasoning on Compliance Monitoring: Balancing Flexibility and Regulation. In Proceedings of the 23rd International Symposium on Foundations of Intelligent Systems, ISMIS 2017, Warsaw, Poland, 23–29 June 2017; pp. 3–16.
41. Schulz, S.; Jansen, L. Formal Ontologies in Biomedical Knowledge Representation. *Yearb. Med. Inform.* **2013**, *22*, 132–146.
42. Cohn, D.; Hull, R. Business Artifacts: A Data-Centric Approach to Modelling Business Operations and Processes. *IEEE Data Eng. Bull.* **2009**, *32*, 3–9.
43. Calvanese, D.; De Giacomo, G.; Montali, M. Foundations of Data-Aware Process Analysis: A Database Theory Perspective. In Proceedings of the 32nd ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2013, New York, NY, USA, 22–27 June 2013; pp. 1–12.
44. Artale, A.; Kovtunova, A.; Montali, M.; van der Aalst, W.M. Modelling and Reasoning over Declarative Data-Aware Processes with Object-Centric Behavioral Constraints. In Proceedings of the 17th International Conference on Business Process Management, BPM 2019, Vienna, Austria, 1–6 September 2019; pp. 139–156.
45. Sun, S.X.; Zeng, Q.; Wang, H. Process-Mining-Based Workflow Model Fragmentation for Distributed Execution. *IEEE Trans. Syst. Man Cybern.* **2010**, *41*, 294–310. [CrossRef]
46. Knoll, D.; Reinhart, G.; Prügmeier, M. Enabling Value Stream Mapping for Internal Logistics using Multidimensional Process Mining. *Expert Syst. Appl.* **2019**, *124*, 130–142. [CrossRef]
47. Loreti, D.; Chesani, F.; Ciampolini, A.; Mello, P. A Distributed Approach to Compliance Monitoring of Business Process Event Streams. *Future Gener. Comput. Syst.* **2018**, *82*, 104–118. [CrossRef]
48. Soueidi, C.; Falcone, Y.; Hallé, S. Monitoring Business Process Compliance across Multiple Executions with Stream Processing. In Proceedings of the 27th International Workshops on Enterprise Design, Operations, and Computing, EDOC 2023, Groningen, The Netherlands, 30 October–3 November 2023; pp. 247–264.
49. Chakraborty, C.; Kishor, A. Real-Time Cloud-Based Patient-Centric Monitoring Using Computational Health Systems. *IEEE Trans. Comput. Soc. Syst.* **2022**, *9*, 1613–1623. [CrossRef]
50. Maitín, A.M.; Nogales, A.; Fernández-Rincón, S.; Aranguren, E.; Cervera-Barba, E.; Denizon-Arranz, S.; Mateos-Rodríguez, A.; García-Tejedor, A.J. Application of Large Language Models in Clinical Record Correction: A Comprehensive Study on Various Retraining Methods. *J. Am. Med. Inform. Assoc.* **2025**, *32*, 341–348. [CrossRef]

51. Kalid, N.; Zaidan, A.A.; Zaidan, B.B.; Salman, O.H.; Hashim, M.; Albahri, O.S.; Albahri, A.S. Based on Real Time Remote Health Monitoring Systems: A New Approach for Prioritization “Large Scales Data” Patients with Chronic Heart Diseases Using Body Sensors and Communication Technology. *J. Med. Syst.* **2018**, *42*, 69.
52. Javed, R.; Abbas, T.; Shahzad, T.; Kanwal, K.; Ramay, S.A.; Khan, M.A.; Ouahada, K. Enhancing Chronic Disease Prediction in IoMT-Enabled Healthcare 5.0 Using Deep Machine Learning: Alzheimer’s Disease as a Case Study. *IEEE Access* **2025**, *13*, 14252–14272.
53. Arora, A.; Chakraborty, P.; Bhatia, M.P.S.; Kumar, A. Deep-SQA: A Deep Learning Model Using Motor Activity Data for Objective Sleep Quality Assessment Assisting Digital Wellness in Healthcare 5.0. *Expert Syst.—J. Knowl. Eng.* **2024**, *41*, e13321. [\[CrossRef\]](#)
54. Raja, R.S. A Multi-Level NLP Framework for Medical Concept Mapping in Healthcare AI Systems. In Proceedings of the 4th IEEE International Conference on AI in Cybersecurity, ICAIC 2025, Houston, TX, USA, 5–7 February 2025; pp. 108–126.
55. Alafari, F.; Driss, M.; Cherif, A. Advances in Natural Language Processing for Healthcare: A Comprehensive Review of Techniques, Applications, and Future Directions. *Comput. Sci. Rev.* **2025**, *56*, 100725. [\[CrossRef\]](#)
56. Liu, Y.; Wang, H.; Zhou, H.; Li, M.; Hou, Y.; Zhou, S.; Wang, F.; Hoetzlein, R.; Zhang, R. A Review of Reinforcement Learning for Natural Language Processing and Applications in Healthcare. *J. Am. Med. Inform. Assoc.* **2024**, *31*, 2379–2393. [\[PubMed\]](#)
57. Baccour, E.; Erbad, A.; Mohamed, A.; Hamdi, M.; Guizani, M. Reinforcement Learning-Based Dynamic Pruning for Distributed Inference via Explainable AI in Healthcare IoT Systems. *Future Gener. Comput. Syst.* **2024**, *155*, 1–17. [\[CrossRef\]](#)
58. Howlader, P.; Pal, K.K.; Cuzzocrea, A.; Kumar, S.D.M. Predicting Facebook-Users’ Personality Based on Status and Linguistic Features via Flexible Regression Analysis Techniques. In Proceedings of the 33rd Annual ACM Symposium on Applied Computing, SAC 2018, Pau, France, 9–13 April 2018; pp. 339–345.
59. Camara, R.C.; Cuzzocrea, A.; Grasso, G.M.; Leung, C.K.; Powell, S.B.; Souza, J.; Tang, B. Fuzzy Logic-Based Data Analytics on Predicting the Effect of Hurricanes on the Stock Market. In Proceedings of the 2018 IEEE International Conference on Fuzzy Systems, FUZZ-IEEE 2018, Rio de Janeiro, Brazil, 8–13 July 2018; pp. 1–8.
60. Leung, C.K.; Braun, P.; Cuzzocrea, A. AI-Based Sensor Information Fusion for Supporting Deep Supervised Learning. *Sensors* **2019**, *19*, 1345. [\[CrossRef\]](#)
61. Austin, C.A.; Mohottige, D.; Sudore, R.L.; Smith, A.K.; Hanson, L.C. Tools to Promote Shared Decision Making in Serious Illness: A Systematic Review. *JAMA Intern. Med.* **2015**, *175*, 1213–1221. [\[CrossRef\]](#)
62. Tran, N.D.T.; Leung, C.K.; Fung, D.L.X.; Mai, T.H.D. Health Analytics on Big COVID-19 Data. In Proceedings of the 2021 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2021, Houston, TX, USA, 9–12 December 2021; p. 1395.
63. Jalil, N.A.; Leen, M.W.E. Big Data in the Era of Pandemic COVID-19: Application of IoT Based Data Analytics, Machine Learning and Artificial Intelligence. In Proceedings of the 5th ACM International Conference on Big Data and Education, ICBDE 2022, Shanghai, China, 26–28 February 2022; pp. 361–367.
64. Arslan, A.; Golgeci, I.; Khan, Z.; Al-Tabbaa, O.; Hurmelinna-Laukkanen, P. Adaptive Learning in Cross-Sector Collaboration During Global Emergency: Conceptual Insights in the Context of COVID-19 Pandemic. *Multinatl. Bus. Rev.* **2021**, *29*, 21–42.
65. Romero, D.; Escudero, P. Adaptive Learning in Agent-Based Models: An Approach for Analyzing Human Behavior in Pandemic Crowding. *Appl. Syst. Innov.* **2023**, *6*, 113. [\[CrossRef\]](#)
66. Cuzzocrea, A.; Belmerabet, I. PROTECTION Pandemic Prevention and Control Dataset. Available online: <https://github.com/ibelmabet-idealab-unical/BDCC2025-datasets> (accessed on 31 December 2024).
67. Croce, F.; Valentini, R.; Maranghi, M.; Grani, G.; Lenzerini, M.; Rosati, R. Ontology-Based Data Preparation in Healthcare: The Case of the AMD-STITCH Project. *SN Comput. Sci.* **2024**, *5*, 437. [\[CrossRef\]](#)
68. Chen, L.; Lu, D.; Zhu, M.; Muzammal, M.; Samuel, O.W.; Huang, G.; Li, W.; Wu, H. OMDP: An Ontology-Based Model for Diagnosis and Treatment of Diabetes Patients in Remote Healthcare Systems. *Int. J. Distrib. Sens. Netw.* **2019**, *15*, 1–15. [\[CrossRef\]](#)
69. Taher, N.C.; Mallat, I.; Agoulmine, N.; El-Mawass, N. An IoT-Cloud Based Solution for Real-Time and Batch Processing of Big Data: Application in Healthcare. In Proceedings of the 3rd IEEE International Conference on Bio-Engineering for Smart Technologies, BioSMART 2019, Paris, France, 24–26 April 2019; pp. 1–8.
70. Addobea, A.A.; Li, Q.; Obiri, I.A.; Hou, J. A Batch Processing Technique for Wearable Health Crowd-Sensing in the Internet of Things. *Cryptography* **2022**, *6*, 33. [\[CrossRef\]](#)
71. Diogo, A. COVID-19 Stats and Mobility Trends Dataset. Available online: <https://www.kaggle.com/datasets/diogoalex/covid19-stats-and-trends> (accessed on 31 December 2024).
72. NYC Health. Coronavirus Data. Available online: <https://github.com/nychealth/coronavirus-data/tree/master> (accessed on 31 December 2024).
73. Haider, S.A. JN.1 (COVID-19 Variant) Sentiment Analysis. Available online: <https://www.kaggle.com/datasets/syedali110/jn-1covid-19-variant-sentiment-analysis> (accessed on 31 December 2024).
74. Hariri, R.H.; Fredericks, E.M.; Bowers, J.M. Uncertainty in Big Data Analytics: Survey, Opportunities, and Challenges. *J. Big Data* **2019**, *6*, 44. [\[CrossRef\]](#)

75. Essaidi, A.; Bellafkih, M.; el Mehdi, K. The Big Data and Machine Learning in Managing Global Health Crises: The Case of the COVID-19 Pandemic. In Proceedings of the 14th IEEE International Conference on Intelligent Systems: Theories and Applications, SITA 2023, Casablanca, Morocco, 22–23 November 2023; pp. 1–8.
76. Demertzis, K.; Taketzis, D.; Tsiotas, D.; Magafas, L.; Iliadis, L.; Kikiras, P. Pandemic Analytics by Advanced Machine Learning for Improved Decision Making of COVID-19 Crisis. *Processes* **2021**, *9*, 1267. [[CrossRef](#)]
77. Masum, M.; Shahriar, H.; Haddad, H.; Faruk, M.J.H.; Valero, M.; Khan, M.A.; Rahman, M.A.; Adnan, M.I.; Cuzzocrea, A. Bayesian Hyperparameter Optimization for Deep Neural Network-Based Network Intrusion Detection. In Proceedings of the 2021 IEEE International Conference on Big Data, BigData 2021, Orlando, FL, USA, 15–18 December 2021; pp. 5413–5419.
78. Faruk, M.J.H.; Shahriar, H.; Valero, M.; Barsha, F.L.; Sobhan, S.; Khan, M.A.; Whitman, M.E.; Cuzzocrea, A.; Lo, D.C.; Rahman, A.; et al. Malware Detection and Prevention using Artificial Intelligence Techniques. In Proceedings of the 2021 IEEE International Conference on Big Data, BigData 2021, Orlando, FL, USA, 15–18 December 2021; pp. 5369–5377.
79. Abad, A.R.K.K.; Barzinpour, F.; Pishvae, M.S. Green and Reliable Medical Device Supply Chain Network Design under Deep Dynamic Uncertainty: A Novel Approach in the Context of COVID-19 Outbreak. *Appl. Soft Comput.* **2024**, *149*, 110964. [[CrossRef](#)]
80. Shiri, M.; Fattahi, P.; Sogandi, F. Two-Stage Approach for COVID-19 Vaccine Supply Chain Network under Uncertainty using the Machine Learning Algorithms: A Case Study. *Eng. Appl. Artif. Intell.* **2024**, *135*, 108837. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.