International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 16 (2024) pp. 469–484

Article

Applying PM4ILP to Identify Loosely or Unstructured Parts in the COVID-19 Patient Treatment Process

Nesrine Missaoui [1,2,*], Maissa Rhif [1] and Sonia Ayachi Ghannouchi [1,2]

1 Higher Institute of Management of Sousse, University of Sousse, Sousse 4000, Tunisia; maissarhif@gmail.com (M.R.); sonia.ayachi.ghannouchi@gmail.com (S.A.G.)
2 RIADI Laboratory, ENSI, University of Mannouba, Mannouba 2010, Tunisia
* Correspondence author: missaouinesrine2015@gmail.com

Received date: 5 March 2024; Accepted date: 21 March 2024; Published online: 10 July 2024

Abstract: This paper presents the application of Process Mining for Identifying Loosely processes (PM4ILP), designed to facilitate the discovery, modeling and improvement of unstructured and/or loosely processes, to the process of treating COVID-19 patients. By following the phases of the cycle and using process mining techniques, we analyzed the behavior of the COVID-19 process, with the aim of improving its effectiveness and efficiency. This was done by collecting real data and applying the phases of the cycle to facilitate the identification of the type of the process and represent improvements on its structure. Our results demonstrate the adaptability and effectiveness of the PM4ILP approach in identifying loosely/unstructured processes and optimizing their quality specifically in critical domains such as healthcare. In addition, we were able to highlight the benefits of PM4ILP, including its ability to facilitate the discovery phase, and the continuous improvement of a process.

Keywords: process improvement; loosely processes; unstructured processes; process mining; COVID-19 process; process quality; PM4ILP lifecycle

1. Introduction

Business process management is an approach that allows organizations to properly manage their business processes by proposing a set of methods and techniques facilitating the modelling, analysis, automation and continuous improvement of processes [1]. It has been widely adopted in both research and industry, as it guarantees consistent results and offers opportunities for managing, execution, and improving business processes [2–4].

The approach presents a lifecycle that is composed of a number of phases allowing the discovery, (re)redesign, enactment, and monitoring of a business process (BP). Within the lifecycle, process discovery and (re)design phases are the most important ones since they allow understanding the way a given process is performed, how it can be modelled and what type of improvement should be applied. In fact, these phases permit determining the structure of the process and verify whether it is adapted to the one presented in the real world or not.

In this context, unstructured processes or loosely processes can be studied. The first are presented as processes that do not have a clear definition on how they are executed or how certain tasks (parts of the process) are performed. They are defined by [5] as “processes that are typically difficult to manage while having a big potential for improvement”. Meanwhile, loosely processes are known as processes that have a defined set of tasks to be executed, where some parts of the process may not be detailed until the time of its execution. They allow a certain degree of flexibility through which process modelling can be completed at run-time.

To be able to discover and improve such processes, process mining techniques can be adopted. These techniques involve extracting knowledge from data generated during the execution of many instances of a process, to better understand the corresponding process and optimize its performance [6]. They allow discovering the behaviour of a process and how it's executed based on the analysis of the event logs.

In literature, process mining techniques were mostly applied on unstructured processes [7,8]. Van Der Aalst considers it as a "backward looking" technique since it allows to study the behaviour of the process on the basis of the event logs generated from the execution traces and to define a model of the process close to what happens in the real world [9]. Nevertheless, it is considered difficult to apply this technique on loosely specified processes. This is because most of these techniques focus on the "structural" aspects of the processes: they do not consider the different changes that a process can have during its execution.

Based on this shortcoming, we proposed in previous research a lifecycle that allows the identification and improvement of unstructured and/or loosely processes [10]. The lifecycle entitled Process Mining for Identifying Loosely processes (PM4ILP) details the phases to follow during the process discovery and modelling phases in order to properly define the type of the process and suggest improvements to its structure.

To validate the feasibility and effectiveness of the PM4ILP approach, we conducted an experimental study in which we applied its phases to the process of treating COVID-19 patients. This field has been extensively studied in literature, where numerous research projects have been undertaken and proposed [11,12]. These studies focused on various aspects of the pandemic, including its impact on public health systems, and treatment options.

Thus, given the dynamic and urgent evolution of healthcare processes including, patient admission, treatment and discharge, this was an ideal environment to test the adaptability and applicability of our approach. Using real data collected from the process, we followed the phases of the approach where we were based on process mining techniques to analyze the COVID-19 treatment process, identify its type and suggest improvement to its structure.

Indeed, two iterations were performed to the COVID-19 process allowing obtaining a more comprehensive and refined process model that better align with real world process. This served to validate the applicability of the approach, enabling us to observe how the PM4ILP lifecycle can be applied effectively to understand and improve rapidly changing processes.

The remainder of the paper is structured as follows. Section 2 presents an overview of the proposed PM4ILP approach. Section 3 details the application of the approach by focusing on the phases of process identification and analysis. The results of process improvement and validation phases are presented in Section 4. Section 5 shows the results of the second improvement iteration. In section 6 a discussion on the feasibility of the approach is presented. Section 6 concludes the paper and underlines some implications for further research.

2. Overview of the Proposed Approach

The approach entitled PM4ILP (Process Mining for Identifying Loosely Processes), defined in previous research [10] allows identifying the type of the process, whether it's loosely or unstructured, and suggesting improvements based on process mining technique. It favors the continuous improvement of a process by allowing it to return from the searching for other improvement insights step to the identification of loosely/unstructured parts step.

It was inspired from the work of [13] where, building upon their work, we have tailored our approach to the challenges of discovering and improving unstructured and loosely specified processes while detailing how process mining can help in applying these phases.

In fact, our PM4ILP is defined as a lifecycle that regroups four main phases: process identification, process analysis, process improvement, and process validation. Each phase has a number of steps conducted to identify and complete the loosely/unstructured parts of a process. Figure 1 provides an overview of the different phases of the approach as well as the inputs and outputs of each phase.

- Process identification provides detailed information about the process domain. It allows understanding the way a BP is performed, i.e., the flow of activities and the involved actors. This allows having a process architecture that represents the structure of a process (called as-is process model).
- Process analysis aims at identifying the type of the process whether it is unstructured or loosely. It is based on four main steps: process discovery, data processing, as-is and discovered process models discovery, and identification of loosely parts. These steps help into providing an analysis on the process model and identifying the loosely/unstructured parts of the process. This could be done by presenting an abstract model as input for the process discovery step and an event log as input for the data processing step.
- Process improvement aims to present a list of improvements that could be applied to the parts

found in the previous phase. Two elements must be provided for this phase. First, the type of the model, either loosely or unstructured. Second, the parts that need improvement. The improved parts represent the output of this phase.

- To achieve this, three actions are required: Identify, select and apply. In other words, identify the types of the improvement, select the most adequate improvement, and apply the selected improvement.
- Process validation permits verifying the quality of the improved process model using a number of quality metrics. If the quality of the model is satisfied then we move to the process implementation step. If not, we return to the step of loosely parts identification from which new improvements can be applied on the process.

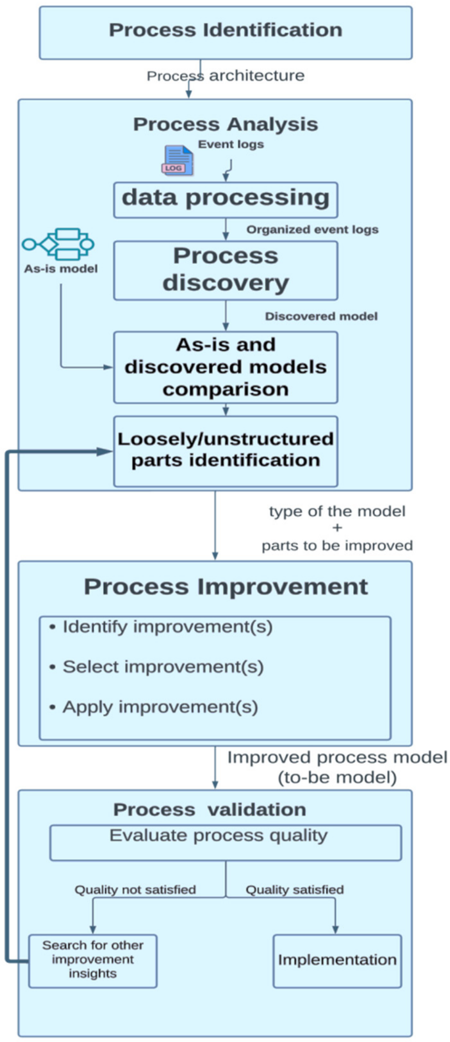

Figure 1. PM4ILP lifecycle phases.

The lifecycle was adopted within a case study where we wanted to improve the given process by determining its type and identifying the parts that need to be improved. The following section shows the details and results of applying the presented phases. In fact, we will first present the results of process identification and analysis phases and then move on to process improvement and validation phases.

3. PM4ILP Application: Process Identification and Analysis

The following sections are devoted to applying the two first phases of the proposed approach on the case study. It aims to provide a detailed application of the process identification and process analysis phases by presenting how each phase was conducted on the COVID-19 process. It will also give a better understanding of the domain selected for this work, the dataset that allowed us to perform the study, and tools used to apply the different phases of the approach.

3.1. Case Study Presentation

We have chosen to work with the healthcare domain which is considered as one of the most important fields where the workflow of executing the tasks of a process cannot be easily determined given the circumstances that the process may adhere to. Thus, it is difficult to predict how the process (or a certain part of the process) is executed.

Indeed, we were primarily interested in the COVID-19 pandemic as a field of study for this work. The COVID-19 lends itself perfectly to this research work where the numerous data that are available in this field and their variety make it suitable for the purpose of this study. As a matter of fact, the COVID-19 process undergoes a number of activities that are performed differently from a case to another (given a patient's health condition) and where the process (parts of it) may change. This in fact reinforces our interest in the field and puts it at a spot of high interest.

We studied how patients are received, examined and treated within Farhat Hached University Hospital Center (UHC), a Tunisian public hospital centralized in Sousse. We worked directly with the infectious disease unit of the hospital that helped us to identify the process workflow and retrieve data needed for the analysis step.

3.2. Process Identification

We focused in this work on the process of treating COVID-19 patients within the infectious department of the hospital where patients are first examined in the emergency department before being transferred and treated in the infectious disease department. Figure 2 presents the COVID-19 process model using Business Process Modelling and Notation (BPMN) standard. It displays the workflow of the actual process, the way it is designed and portrays how the activities of the process are connected together.

For the infectious disease department, the process starts with the task of receiving a patient. Based on the results of the examination, two cases are performed: either to exclude the patient if no signs of COVID-19 have been presented or to simultaneously conduct three tasks "Start isolation", "Assess patient health status" and "Conduct a PCR test". Then status of a patient will be checked, and three activities will be performed: either the "Monitor patient at home", or "Prepare hospitalization" or "Transfer patient to intensive care unit".

If the task "Prepare hospitalization" is performed, then another decision will be made. If the patient is new then two possible activities will be performed: "Lift isolation" if COVID is not confirmed and "Keep isolation" if COVID is confirmed. In case the disease is confirmed, then a task of choosing COVID-19 procedure is performed. Right after this task, three activities will be presented: "Monitor at home" if the patient's health status is improved, "Transfer to containment center" if it requires further monitoring, or "Declare time of death" in case of death.

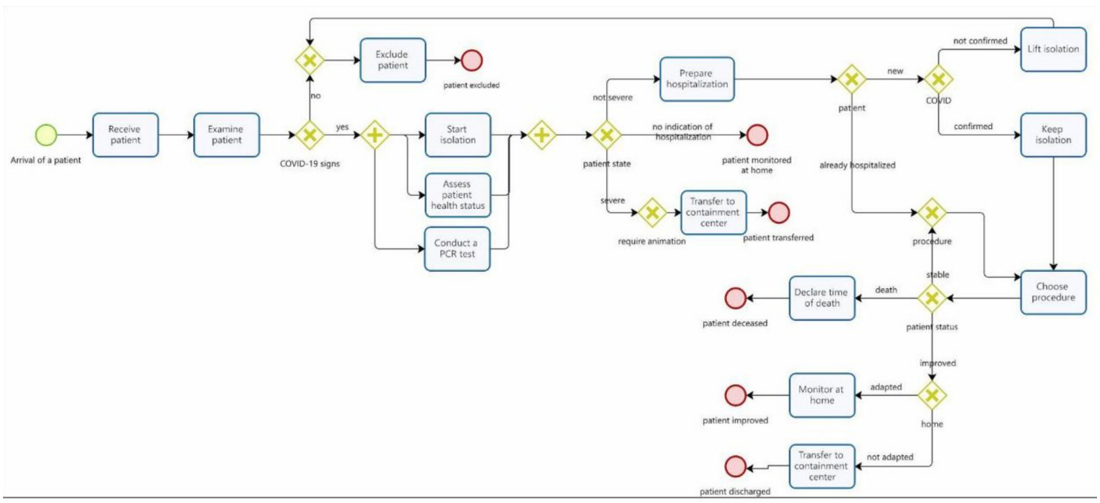

Figure 2. As-is COVID-19 process model.

3.3. Process Analysis

After identifying the process model, we moved to the process analysis where we followed the steps presented in the cycle to identify the type of the process and the improvement that can be performed on the process.

Within this phase: three steps were conducted: data processing, as-is and discovered models comparison, and loosely/unstructured parts identification.

3.3.1. Data Processing and Process Discovery Steps

The data used for this case study is composed of 2095 lines of both positive and negative COVID 19 cases; it contains cases treated from March 2019 to June 2019. The output of this step is an organized Excel spreadsheet that contains only relevant and valuable data. It starts with obtaining an Excel file, then checking the data, and filling in the missing data for certain patient attributes through visits to the infectious diseases department. After the steps of data filtering and processing, we were able to obtain an event log regrouping the different cases. Each case is presented by the following attributes: entry mode, task, start date, end date, actors and COVID state. Figure 3 presents an excerpt of the generated event log.

Case id	Task	Start date	End date	Actor	Covid
1p	Receive patient	3/3/20 8:00	3/3/20 8:10	interne	positive
1p	Examine patient	3/3/20 10:00	3/3/20 10:30	interne	positive
1p	Start isolation	3/3/20 11:30	3/3/20 11:40	interne	positive
1p	Assess patient health status	3/3/20 11:45	3/3/20 11:50	interne	positive
1p	Conduct a PCR test	3/3/20 12:00	3/3/20 12:05	interne	positive
1p	Assess severity	3/3/20 12:10	3/3/20 12:15	assistant	positive
1p	Orient toward infectious disease unit	3/3/20 12:20	3/3/20 12:25	head of the service	positive
1p	Decide on the diagnostic/procedure	3/3/20 12:30	3/3/20 12:43	professor	positive
1p	Get PCR result	3/3/20 12:55	3/3/20 13:05	Head of the service	positive
1p	Do a clinical examination	3/3/20 13:10	3/3/20 13:16	interne	positive
1p	Check disease presence	3/3/20 13:21	3/3/20 13:26	interne	positive
1p	Keep isolation	3/3/20 13:32	3/3/20 13:36	professor	positive
1p	Choose COVID procedure	3/3/20 13:46	3/3/20 13:56	Associate physician	positive
1p	Start COVID treatment	3/3/20 14:03	3/3/20 14:11	nurse	positive
1p	Start COVID treatment	3/3/20 14:16	3/3/20 14:22	nurse	positive
1p	Start COVID treatment	3/3/20 14:28	3/3/20 14:38	nurse	positive
1p	Conduct a medical check-up	3/3/20 14:58	9/3/20 15:08	nurse	positive
1p	Control patient evolution after treatement	9/3/20 15:17	9/3/20 15:46	Head of the service	positive
1p	Choose COVID procedure	9/3/20 15:50	9/3/20 15:59	Associate physician	positive
1p	Conduct a medical check-up	9/3/20 16:18	9/3/20 16:28	nurse	positive
1p	Conduct a medical check-up	9/3/20 16:37	9/3/20 16:57	nurse	positive
1p	Do a PCR test	9/3/20 17:07	9/3/20 17:14	professor	positive
1p	Control patient evolution after treatement	15/3/20 14:04	15/3/20 14:10	Associate physician	positive
1p	Choose COVID procedure	15/3/20 14:20	15/3/20 14:30	professor	positive

Figure 3. COVID-19 event log.

3.3.2. As-is and Discovered Models Comparison

Following the data processing step, the discovery of the process model is performed where we relied on the created event log to generate the process model. This step was conducted using process mining tools from which we were able to analyze and discover our model. Indeed, two tools were adopted: ProM and BPMNDiFFViz. ProM was used for the fact that it offers numerous algorithms to discover a process model. It also allows an in-depth analysis of the performance of a process using a number of metrics [14]. BPMNDiffViz was applied for its ability to understand, analyze, and manage BPMN diagrams. It visualizes and compares business process model and notation (BPMN) diagrams. In fact, BPMNDiffViz serves as an easy comparison by highlighting the differences of two diagrams side by side. It also allows a workflow analysis, where we can analyze and understand how modifications affect the sequence of activities [15].

For ProM, we have chosen to work with two discovery algorithms: Fuzzy miner and inductive miner. Fuzzy miner is an algorithm that permits to mine unstructured or loosely processes by discovering and highlighting the missing activities and presenting all the possible paths that could be found in a model. The result of the algorithm is defined in a Petri net model where the type of gateways is not well presented, i.e., it does not explain whether the activities are performed exclusively, inclusively, or simultaneously. To solve this problem, we resorted to inductive miner algorithm which is introduced by [16] as a solution that supports incompleteness reproduced by the fuzzy miner. It allows discovering a process and converting the model to a BPMN diagram that distinguishes correctly AND relations and XOR relations [6].

Thus, after choosing the appropriate algorithms, we began with the discovery of the COVID-19 process using fuzzy miner. The application of the algorithm produced a mined process model (called also Fuzzy model) that displays the way activities are performed according to the event log. Figure 4 shows an excerpt of the discovered fuzzy model from the imported event log. This model follows a number of activities and suggests a bunch of discovered activities such as “choose COVID procedure”, “Do an imaging check-up” and so many others. Each activity is presented with the values of two parameters: correlation and significance. The significance parameter determines the degree of interest in events or their occurrence after one another by assessing their frequency while correlation calculates the degree of closeness between two consecutive events.

These two parameters were used to study the behaviour of the process and determine the relation between the activities and their importance. Indeed, we mainly focused on identifying the most frequently performed activities in the process. To do this, we relied on the thickness and darkness of the edges linking two activities, i.e. the darker and thicker the edge, the higher its significance and correlation values. We take as an example; the edge connecting the two activities “Choose COVID procedure” and “Control patient evolution after treatment”, the thickness of the edge validates the importance and connection of these two activities.

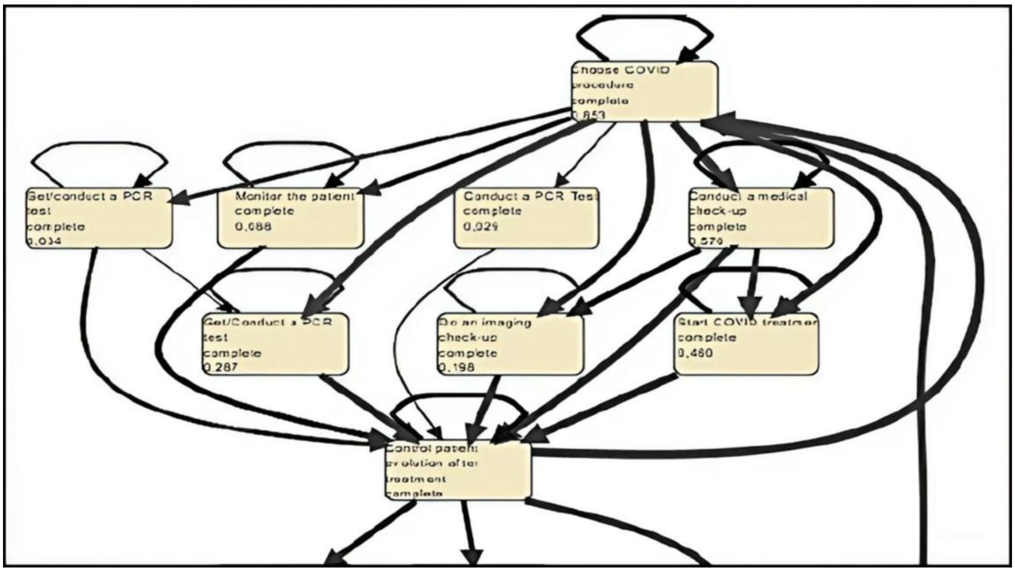

Figure 4. Discovered fuzzy model.

To identify the type of gateways (parallel, inclusive, exclusive), we applied the inductive miner algorithm where we first converted the event log to an XES format. After converting the log, we selected the value of the “Noise threshold” which is equal to 0.2. This precise value guarantees a not too detailed and complex model where it only presents the most relevant paths. We were able, through this algorithm, to obtain a BPMN diagram that shows the workflow of activities and their exact connections. Figure 5 presents an excerpt of the BPMN diagram.

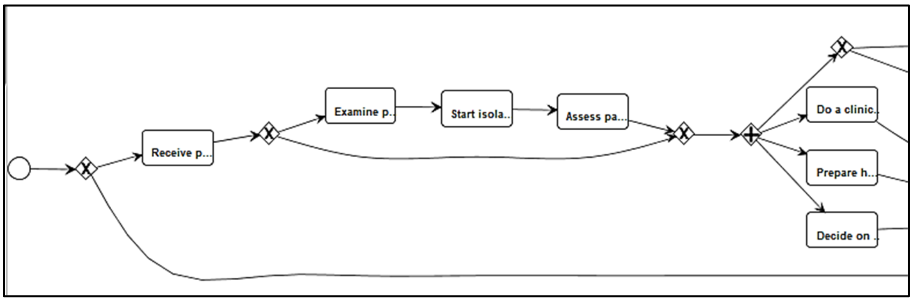

Figure 5. Discovered BPMN diagram using inductive algorithm.

3.3.3. Loosely/Unstructured Parts Identification

The main goal of this step is to identify the type of the process (loosely and/or unstructured) and determine the parts that need improvement. To do so, we compared the real world model (as-is model) with the discovered model. This comparison takes into consideration the existence of the performed activity in the two models (present in both models or in one of them), its placement (where is it placed in the two models) as well as the control flows of the two models (the way an activity is connected to another).

In this step, we conducted two different types of comparison, first we compared the as-is model with the fuzzy model, and then we compared the as-is with the model discovered from the inductive miner algorithm. Fuzzy miner comparison shows all the activities presented in the given event log and provides the user with the ability to configure the value of displaying all the possible edges. On the other hand,

the inductive miner highlights how the activities are connected and shows the missing gateways.
By comparing the as-is model with the fuzzy model, we were able to obtain the following results:

- Missing activities after the activity "Choose COVID procedure": five activities are performed directly after choosing the procedure which are not presented in the as-is model (start COVID treatment, conduct medical check-up, get/conduct a PCR test, do an imaging exam, monitor the patient)
- An activity of controlling the evolution of the patient is performed right after the missing activities.
- In the fuzzy model, a new activity of examining home conditions is performed after controlling the evolution of the patient from which the patient can be monitored at home.
- Missing activity in the as-is model: an activity of observing the state of the patient takes place just before the execution of the task "declare time of death".
- The fragment related to preparing the hospitalization presented in the as-is model lacks other tasks derived from the fuzzy model. These tasks are performed between the fragment "prepare for hospitalization" and" lift isolation".

From what we found in the Fuzzy model, we can note that the two models are not following the same paths: some tasks do not exist in the as-is model, the position of these tasks in the fuzzy model are not exactly the same as the one presented in the as-is. Hence, we can conclude that the type of the COVID-19 model is loosely since all the compared subparts of the model are identified as loosely.

The results of the comparison allowed determining the real behaviour of the process and the order to which activities are performed. However, it was difficult to determine the type of flow between the activities. Thus, we moved to conducting a comparison of the as-is model and the inductive model using the BPMNDiffViz tool. We began by exporting the inductive model to a BPMN file and then we entered the two models to the tool. As shown in Figure 6, the results of the comparison are visualized using three types of colors where each color corresponds to elements that refer to a particular suggestion:

- The blue elements refer to the matching elements between the as-is model and the discovered model.
- The elements that are marked with the red color are the elements that BPMNDiffViz suggests to delete.
- The elements that are marked with the green color are the elements that BPMNDiffViz suggest to add.

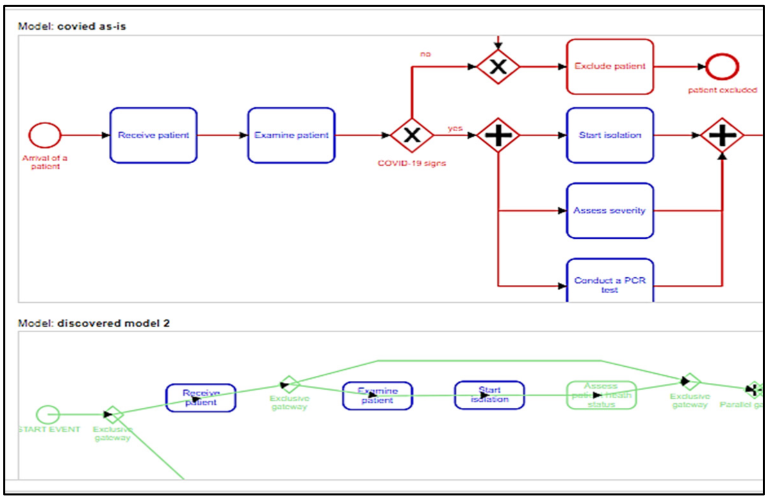

Figure 6. Results of the comparison using BPMNDiffViz.

Based on the results obtained from both comparisons, we studied the type of improvements that could be added to the COVID-19 process model where we concentrated on the common parts that were deduced from the fuzzy miner and the BPMNDiffViZ. Table 1 shows a list of common suggestions identified from the two tools.

Table 1. Common results and suggestions between Fuzzy and BPMNDiffViZ.

Fuzzy Miner Comparison	BPMNDiffViZ Comparison
In the fuzzy model result, “Observe the state of the patient” takes place just before the execution of the “Declare time of death”.	BPMNDiffViZ also shows that “Observe the state of the patient” should take place just before the execution of the “Declare time of death”.
In the Fuzzy model, “Control patient evolution after treatment” task is placed after the execution of the two tasks “Choose COVID treatment” and “Start COVID treatment”.	BPMNDiffViZ shows the same result also, and the same placements of the task “Control patient evolution after treatment”.
In the fuzzy model, “Assess patient health status” is performed before the task “conduct a PCR test” and right after “Start isolation”.	BPMNDiffViZ also suggests the same placement for the later task.
The Fuzzy model shows that the tasks: “decide on the diagnosis/procedure", "do a clinical examination”, “get PCR result" and "check disease presence" were performed after the task “Start isolation” and before “Keep isolation” task.	The result of BPMNDiffViZ shows the same sequence of activities that the fuzzy suggested.

As a sum up, the comparison results finalizes the analysis phase from which we were able to determine the type of the model and the parts that need to be improved. Having detected the outputs of this phase, we are now able to move to the next phases which are “process improvement” and “process validation”.

4. Process Improvement and Validation

The outputs of the analysis phase are considered in the improvement phase where three actions are realized: identify improvements, select the adequate improvement and apply improvements.

4.1. Process Improvement

4.1.1. Identify improvements

Based on the results obtained from the BPMNDiffViZ tool, we were able to highlight the missing parts (activities, paths, and gateways) and propose the right and exact placement for them. The improvements are classified into two categories: addition and deletion.

Table 2 details the list of 13 suggested improvements classified in the categories addition and deletion. These improvements were deduced from both fuzzy miner and BPMNDiffViZ.

Table 2. Categories of suggested improvements.

Category		Suggested Improvement
Addition	1	Add a XOR gateway and the activity “Observe the state of the patient” just after the start point and connect it with “Declare time of death” activity.
	2	Add an OR fragment connecting the activities “Do a clinical examination”, “Decide on the diagnosis”, “get PCR results" and "check disease presence" between the two activities “Prepare hospitalization” and “lift isolation”.
	3	Add the activity “Assess patient health status” before “Prepare hospitalization” and after “Start isolation”. The activity is located between the two gates “Patient status” and “COVID” gateways.
	4	Add a new fragment composed of five activities after the activity “choose COVID procedure”..
	5	Add the activity “control evolution after treatment” before the gateway “patient status” and link it to the new fragment.
	6	Add the activity “examine home condition” after the gateway ‘patient status” (case of healed patient).
Deletion	7	Delete the path between the start point and “Receive patient” activity.
	8	Delete the path between the activities “Receive patient” and “Examine patient”.
	9	Delete the entire path between the gateways “COVID-19 signs” and “patient state”.
	10	Delete the paths between the gateways “Patient state” and “Patient”.
	11	Delete the activities “Lift isolation” and “Choose COVID-19 procedure”.
	12	Delete the paths from the gateway “Procedure” until “Home” gateway.
	13	Delete the fragment that comes right after the gateway “home”.

Having concluded the various types of improvements suggested to be applied, we moved to the next step which is selecting the appropriate improvement.

4.1.2. Select Improvements

The selection of the appropriate improvements to apply required the inclusion of the medical staff in order to appropriately select the relevant improvements. Table 3 synthesizes the different improvements deduced from the tool together with the opinion of medical staff. It explains and demonstrates the list of improvements selected. These improvements are classified into two distinct categories: Addition and deletion.

Table 3. Selected improvements.

Improvement Category	N°	Valid	Invalid	Justification
Addition	1	✔		This was selected by the medical staff as an important modification that should be applied.
	2	✔		The medical staff chose the improvement, concluded from the fuzzy and the tool as a logical solution.
	3	✔		The medical staff approved the addition of the activity.
	4	✔		The medical staff found that the result concluded by the tool is more logical.
	5	✔		This improvement was selected by the medical staff and will be added to the to-be model.
	6		✔	This improvement was not selected by the medical staff, because according to them the missing activity was not important.
Deletion	7	✔		This improvement is approved, because the deleted path will be replaced with XOR gateway that will relate the activity “Receive patient” with the start event with a supplementary path.
	8		✔	Improvement is not selected, because this path is important and shouldn’t get deleted. Without it the model will be presented in an illogical way.
	9		✔	The two gateways “COVID-19 signs” and “Patient state” are important and represent relevant elements. However, these gateways should be presented in the new model.
	10		✔	The mentioned elements are important and shouldn’t get deleted, even the paths are correct, that’s why Improvement n°4 is not valid.
	11		✔	The medical staff did not validate improvement n°6, because both of the mentioned activities are important and shouldn’t get deleted.
	12	✔		The mentioned paths should get deleted and replaced with other discovered activities, gateways and new paths.
	13		✔	According to the medical staff, the elements marked in red are relevant and should be presented in the improved model.

4.1.2. Apply Improvements

Having selected the appropriate improvements with the help of the medical staff, the next step is to apply them to the model. We applied the selected improvements on the COVID-19 process model from which we were able to obtain an improved version of the process (the to-be model). Figure 7 illustrates the improved model that only includes the relevant activities, paths and gateways selected by the medical staff. The elements which are marked in green represent the new elements that we added and the improved parts.

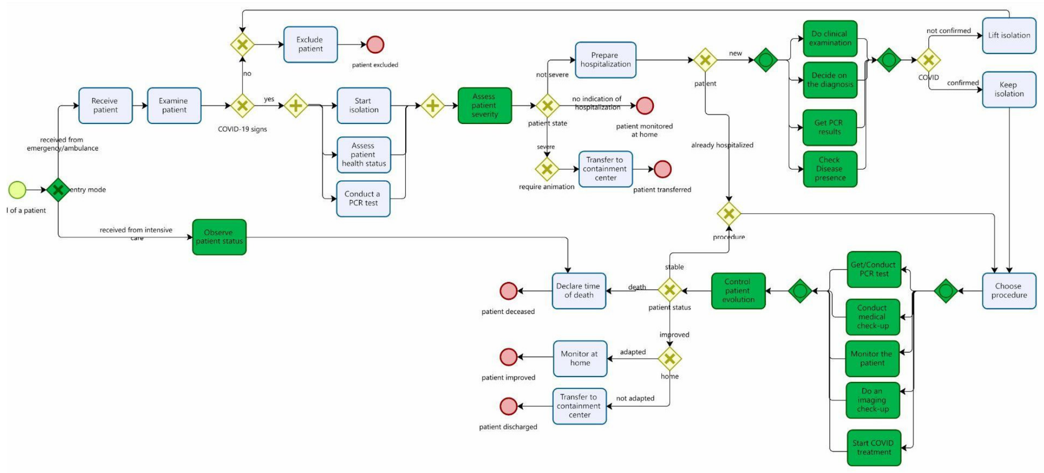

Figure 7. Improved "to-be" process model.

4.2. Process Validation

To validate the efficiency of the PM4ILP cycle and verify the completeness of the process model, two stages of validation were conducted: user satisfaction and process quality validation. The first allows verifying the satisfaction of the medical staff regarding the obtained to-be process. The second permits checking the quality of the improved process by comparing it to the as-is process. This is based on a number of model quality metrics that were selected to validate if the obtained to-be model can be considered as a structured model or not.

For user satisfaction validation, feedbacks were collected from the medical staff to assess the effectiveness of the given enhanced model. They were able to express their satisfaction or dissatisfaction with the current business process model, as well as suggest areas for further improvement.

As for process quality validation, we adopted a number of quality metrics that were proposed by [17]. The authors presented a Lindland's framework that regroups three categories of quality: syntactic quality, semantic quality and pragmatic quality.

- Syntactic analysis focuses on the structural components of a model, with the goal of ensuring that all statements comply with the syntax of the language. To measure this metric, we focused on two criteria: size and complexity. The size criterion measures the number of activities in a process [18], while the complexity criterion evaluates the presence of different splits (XOR, OR, AND) by calculating the Control Flow Complexity (CFC) value [19].
- The CFC value is calculated according to the number of gateways used in the model. For AND split, the CFC value increases by 1 for each split. For XOR split, n possible states are considered for all presented paths thus the value of the CFC metric by n. For an OR partition with n outgoing transitions, there are 2n−1 ways to handle the n outgoing transitions. Therefore, each OR split with n outgoing transitions adds 2n−1 to the CFC metric [19].
- Semantic quality is used to enhance the correspondence between a model and a domain. We chose two criteria: Validity and Completeness. Validity is related to the accuracy of all statements put forward by the model to solve the problem [17], while completeness means that all necessary relevant statements about the problem domain are included. Both of these two criteria are related where each one of them depends on the other.
- Pragmatic quality refers to the consumer perspective [20], precisely in this case the medical staff is in charge of validating the improved model, by judging their flexibility, and understandability. Flexible models can handle variations with the presence of alternative paths to handle different scenarios. While understandable models have an efficient organization of the sequence of activities.

Based on these metrics, we analyzed the quality of the obtained model and compared it to the initial model (as-is). Table 4 presents the results of the comparison using the defined quality criteria where we used the following values:

- (-) Means that the criterion is not satisfied.
- (±) Means that the criterion is partly satisfied and the model still requires improvements.
- (+) Means that the criterion is satisfied

Table 4. Validation results.

Model Qualities		As-is Model	Improved Model
Syntactic	Size	14	26
	Complexity	15	63
Semantic	Validity	-	±
	Completeness	-	±
Pragmatic	Flexibility	±	+
	Understandability	-	±

Based on the results presented in the table above, several interpretations are presented below:

- The size of the improved model is equal to 26, compared to 14 for the as-is model. The higher value of the to-be model is explained by the fact that the improved model includes all the relevant elements to properly treat a patient.
- After applying the CFC formula, we came out with a value equal to 63. In our case, it is predictable and understandable to find a higher CFC value because of the addition of the missing parts. But, despite its higher value, this model is better organized, better structured and more efficient than the initial model that has a CFC value of only 15, but lacks particular relevant elements.
- In our case, the medical staffs is responsible of checking the validity criterion, where they state that all the possible important combinations of the state variables of the domain are included. The dependency of the new model was also checked, in a manner where there is no single state of the new model which needs or waits for external events to happen. But at the same time they ensure that the sequence of activities, decision points, and information flows are not all of them meaningfully and logically connected. According to the medical staff, the model still needs more improvements to be validated.
- In our case, all the process components are identified and presented such as the activities, the decisions and the paths. We have also ensured that all the necessary inputs and outputs are included and properly presented. But not all the missing parts are replaced in the right place. Based on these results, we can conclude that the improved process model still needs more improvements because the organization and the structure of the process are not fully completed. This was also deduced from the above-mentioned criterion "validity of the model".
- The medical staff considered the improved model as flexible because it presents all the different necessary scenarios and paths. While the as-is model was flexible in a way where data can be modified and customized, but it does not cover all the different scenarios.
- The medical staffs find the improved model pretty much understandable, because most of the activities are organized in an understandable sequence. But there is some part of the process that still needs more organization of its activities ("Decide on the diagnosis" and "Check disease presence"). That's why a second iteration will be done to the model so that it can be very understandable to the end users.

As a sum up, the medical staff determined that further changes are necessary where they identified that the activities "Decide on the diagnosis" and "Check disease presence" are not appropriately placed within the sequence, thus lacking logical and meaningful flow.

Based on their assessment, these activities need to be repositioned. "Decide on the diagnosis" should be performed just right after the gateway "Patient". "Check disease presence" activity should be performed after "Do clinical examination" and "Get PCR results" activities and before the "COVID" gateway.

Additionally, the medical staff recommends removing the "exclude patient" activity, the gateway "Home", and the activity "Transfer to containment center" as they are deemed less important, according to the results obtained by BPMNDiffViZ. They also recommend keeping "Monitor at home" activity. By updating these elements, only the most relevant components will be executed in the new model. Consequently, a second iteration is performed, which involves incorporating the feedback from the previous steps to further enhance the process.

5. Second Iteration Results

Within this iteration, we begin directly with the step of identifying loosely or unstructured parts given the fact that the list of improvements has been already identified by the staff members. We will detail in the following sections the results of these steps and the obtained to-be model.

5.1. Identify Loosely/Unstructured Parts of the Model

Based on the last output of the first iteration and according to the medical staff, we were able to deduce the loosely parts of the model. As presented in Figure 8, all the elements that are circled in red represent the loosely points of the model, and these points indicate the areas that require further improvement.

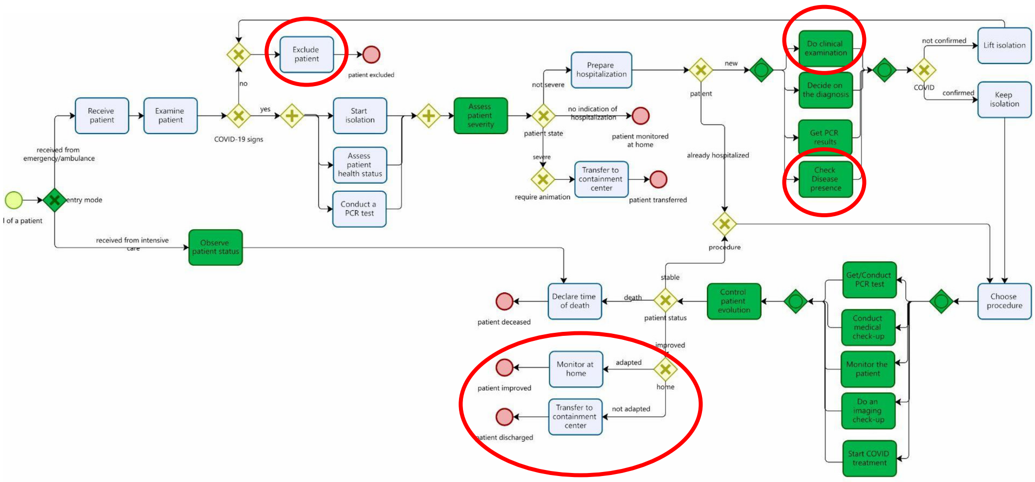

Figure 8. Loosely points identification: second improvement iteration.

5.2. Identify Improvements

The enhancements in this iteration can be categorized into two groups: the "update category" and the “deletion Category". All the improvements that were applied in this iteration are presented in Table 5.

Table 5. Improvement categories for the second iteration.

Deletion Category	**Update Category**
Delete “Exclude Patient” activity.	Replace “Decide on the diagnosis” activity. This activity will be performed right after “Patient” gateway.
Delete “Home” Gateway.	Replace “Check disease presence” activity. This activity will be performed right before “COVID” gateway.
Delete “Transfer to containment center” activity.	Update “Monitor at home” activity placement. This activity will be added to the “Patient status” gateway, where it represents the case of cured patients.

5.3. Select and Apply Improvements

Based on the identified improvement list, we were able to obtain an improved version of the COVID-19 process where modifications related to the placement of a number of tasks were performed in certain parts of the process. Figure 9 shows the new model, which only includes the necessary activities, pathways, and gateways based on the medical staff feedbacks. The green elements represent the modified parts.

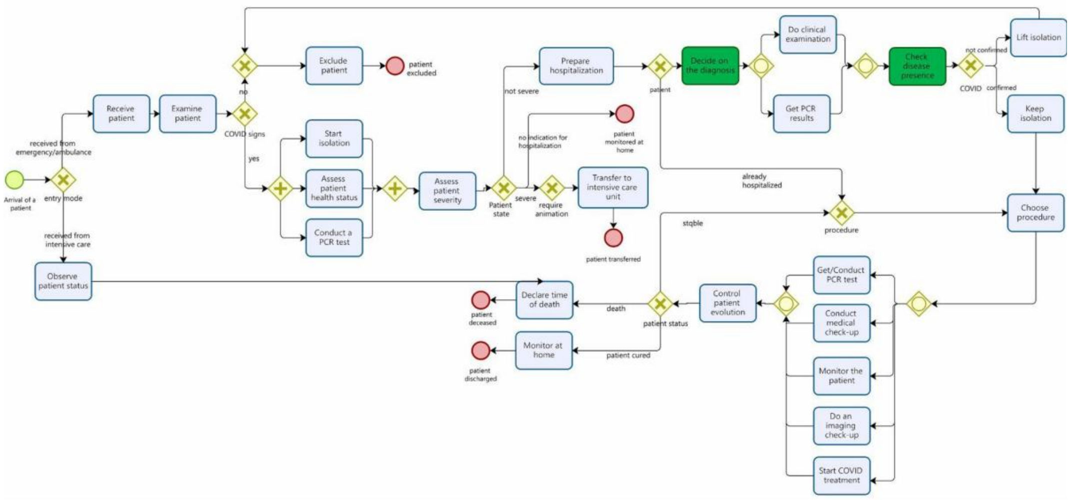

Figure 9. Improved version of the COVID-19 model.

After obtaining the improved process, we proceeded with a validation of the quality of the model where we compared the quality of the three models: as-is model, first to-be and second to-be model using the identified quality criteria. Table 6 presents the results of the comparison.

Table 6. Validation results of the second iteration.

Model Qualities		**As-is Model**	**Improved Model (First Iteration)**	**Improved Model (Second Iteration)**
Syntactic	Size	14	26	23
	Complexity	15	63	49
Semantic	Validity	-	±	+
	Completeness	-	±	+
Pragmatic	Flexibility	±	+	+
	Understandablity	-	±	+

In the second iteration, significant improvements have been made, reflected in both the size and efficiency of the model. The size of the second improved model has been optimized to 23, compared to 26 in the initial iteration.

This reduction underlines the enhanced efficiency of our solution, since the improved model now encompasses all relevant elements. Moreover, this improvement has led to a decrease in complexity, demonstrated by the reduction of the CFC value from 63 to 49. This allowed enhancing the model comprehensibility and organization, making it more structured and coherent.

Medical staff members have validated the model and considered it ready for implementation. Indeed, following detailed iterations, all missing components have been integrated into the business process model, ensuring its completeness. In particular, the improved model in this iteration is praised for its increased clarity and organization, with activities presented in a logical sequence without persistent organizational deficiencies.

6. Discussion

The application of the PM4ILP approach to the identification of unstructured and loosely processes, particularly within the COVID-19 patient treatment process, have demonstrated several benefits and outcomes. Our study showcased the adaptability and efficiency of the proposed lifecycle in identifying the type of the process and understanding what type of improvements to conduct and how to apply them. One of the primary benefits of PM4ILP lies in its systematic and iterative nature, which allows for the continuous enhancement of process models. Through two iterative cycles, we were able to improve the COVID-19 process, resulting in a more complete and accurate representation of the COVID-19 treatment process.

In literature, the concepts of loosely and/or unstructured process improvement were highly studied. The work proposed by [8] aimed to discover process models from incomplete log using the fuzzy miner algorithm. The authors aimed to find solutions to the complexity and variability of data found in processes by attempting to improve unstructured parts within the process. Contrary to this work, [16] suggested building a process model rather than discovering it, which implies modifying the entire process. The authors proposed a modeling approach that adopts process mining techniques to model unstructured processes using more than one type of algorithm, including Fuzzy algorithm, Inductive and Heuristic algorithm.

Improvement of unstructured business processes at run-time was studied by [13]. The work focused on the re-design process where the author concentrated on changing fragments of a process at run time rather than the whole process while being based on process mining techniques. This allowed improving a business process by undergoing various changes on its activities without completely changing its structure. However, no details regarding the tools used or even the results of the case study were clearly explained in the research.

As for loosely process improvement, to the best of our knowledge, few researches were proposed in this context where [21] proposed DECLARE, an open-source workflow management system, inspired by the well-known workflow patterns. The authors stated that the tool allows the identification of structured, unstructured and loosely processes but did not specify the way looseness is treated in the given system. Only the work of [5] studied this aspect by proposing a taxonomy of decision deferral that defines criteria to identify loosely processes using the concept of patterns. This taxonomy explains the different means that can be adopted and illustrates the different patterns that can be used within loosely specified processes.

Comparing our approach to existing literature revealed distinct advantages. From what we found, the majority of proposed studies focus on unstructured process identification and improvement with little details on the steps followed. In addition, loosely processes were hardly studied in literature with no clear explanation on the way process mining is applied during the phases of process discovery and improvement.

PM4ILP lifecycle addresses these limitations by providing a systematic framework for identifying and improving unstructured and loosely defined processes. Moreover, the case study demonstrates the practical applicability of the lifecycle in improving COVID-19 process efficiency. By identifying the type of the process, and the parts to be improved, we were able to propose useful recommendations for enhancing the quality and efficiency of the process. These findings highlight the potential of our PM4ILP approach to deliver significant and appropriate improvements in both unstructured and loosely processes, ultimately leading to better process quality and performance.

7. Conclusions

The goal of this paper is to demonstrate the applicability of the PM4ILP approach by presenting the results of a case study related to the improvement of the process of treating COVID-19 patients. It gives detailed instructions on the way the lifecycle was adopted to (1) discover the behaviour of the process using process mining technique, (2) determine its type and the loosely/unstructured parts, (3) suggest and apply improvements, and (4) validate the effectiveness of the applied improvements.

The application of the lifecycle to the COVID-19 patient treatment process has yielded promising results and demonstrated its efficacy in enhancing process quality. Through iterative improvement and application of process mining techniques, we have successfully identified and improved the quality of the COVID-19 patient's treatment and proposed a process model that meets the needs of the medical staff.

Indeed, the results highlighted the adaptability and applicability of the PM4ILP lifecycle, demonstrating its ability to enhance unstructured or loosely processes using process mining analysis. This empirical validation not only reinforces the theoretical foundations of the PM4ILP approach, but also paves the way for its practical adoption and implementation in a variety of organizational contexts.

For future work, we could explore other applications of the PM4ILP cycle in different healthcare contexts and domains beyond COVID-19 treatment. It would also be useful to focus on validating the applicability of the cycle at runtime, and to investigate the automatic identification and discovery of structured or unstructured parts. This will help extend the cycle's capabilities to facilitate process discovery, modelling and improvement.

Author Contributions

N.M., M.R. and S.A.G. conceived this research and designed experiments; N.M. and S.A.G. participated in the acquisition of data; M.R. performed experiments and analysis; N.M. participated in the analysis and interpretation of the data; N.M. and M.R. wrote the paper and participated in the revisions of it; S.A.G. validated the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding
This research received no external funding.

Conflict of Interest Statement
The authors declare no conflict of interest.

Data Availability Statement
Not applicable.

References

1. Dumas, M., Rosa, L. M., Mendling, J., & Reijers, A. H. *Fundamentals of business Process Management*. Springer-Verlag, 2018.
2. Ellouze, F., Chaâbane, M. A., Andonoff, E., & Bouaziz, R. On the Automation of Managing and Retrieving Versions of Subsea Pipeline Installation Process. *International Journal of Computer Information Systems and Industrial Management Applications*, Vol *10*, pp.238-251, 2018
3. López-Pintado, O., Dumas, M., Berx, J. Discovery, simulation, and optimization of business processes with differentiated resources. *Information Systems*, 120(5), 2024
4. Sinatriya, I. R., Hanggara, B. T., & Rachmadi, A. Evaluation and Improvement of Business Processes in the Operational Division Using Business Process Improvement (BPI) *(Case Study: PT. Mitrasukses Engineering Indonesia. MATICS: Jurnal Ilmu Komputer dan Teknologi Informasi*, *Journal of Computer Science and Information Technology)*, 16(1), pp.30-35, 2024.
5. Reichert, M., Weber, B. Process-aware information systems. Enabling Flexibility. In *Proceedings of the Process-Aware Information Systems: Challenges, Methods, Technologies*, 9-42, 2012.
6. Van Der Aalst, W. Process mining: Data science in action, Vol. 2, Heidelberg, Springer, 2016.
7. Folino, F., Greco, G., Guzzo, A., Pontieri, L. Discovering multi- perspective process models: The case of loosely-structured processes. In *Enterprise Information Systems: 10th International Conference*, *ICEIS, Barcelona, Spain, June 12-16,* pp. 130-143, 2008.
8. Zakarija, I., Skopljanac-Macina, F., Blaskovic, B. Discovering pro-cess model from incomplete log using process mining. In *Proceedings of the 57th International Symposium ELMAR (ELMAR)*, pp.117-120, IEEE, 2015.
9. Van der Aalst, W. M. Process mining and simulation: A match made in heaven!. In SummerSim, pp. 4-1, 2018.
10. Rhif, M., Missaoui, N., Ayachi, G.A. PM4ILP: An approach for the identification and improvement of unstructured and loosely specified processes. In *Proceedings of the 23rd International Conference on Intelligent Systems Design and Applications*, December 11-13, 2023.
11. Bharati, S., Podder, P., Mondal, M., & Prasath, V. B. Medical imaging with deep learning for COVID-19 diagnosis: a comprehensive review. *International Journal of Computer Information Systems and Industrial Management Application*, Vol 13, pp. 91-112, 2021.
12. Bellouch, A., Boujnoui, A., Zaaloul, A., & Haqiq, A. Three-dimensional Markov chain model to help reduce the spread of COVID-19 in IoT environment. *International Journal of Computer Information Systems and Industrial Management Applications*, Vol 13, pp.275-285, 2021.
13. Lamghari, Z. Unstructured Business Processes Improvement using Process Mining Techniques, *ASM Science Journal*, 17 (1), 2022.
14. Van der Aalst, W. M., van Dongen, B. F., Günther, C. W., Rozinat, A., Verbeek, H. M. W., & Weijters, A. J. M. M. Prom: The process mining toolkit. In *Proceedings of the BPM 2009 Demonstration Track (BPM Demos 2009*, Ulm, Germany, September 8, 2009), pp. 1-4, CEUR-WS.org, 2009.
15. Ivanov, S. Y., Kalenkova, A. A., & Van der Aalst, W. M. BPMNDiffViz: a tool for BPMN models comparison, BPMcenter.org, 2015.
16. Stefanini, A., Aloini, D., Benevento, E., Dulmin, R., Mininno, V. A process mining methodology for modeling unstructured processes. *Knowledge and Process Management*, 27(4), pp. 294-310, 2020.
17. Moody, D. L., Sindre, G., Brasethvik, T., Sølvberg, A. Evaluating the quality of process models: Empirical testing of a quality framework. In *Proceedings of the Conceptual Modeling—ER 2002: 21st International Conference on Conceptual Modeling Tampere*, Finland, October 7–11, pp. 380-396, Springer, 2003.
18. Van Belle, J. P. A proposed framework for the analysis and evaluation of business models. In *Proceedings of the 2004 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists on IT Research in Developing Countries*, pp. 210-215, 2004.
19. Cardoso, J. Control-flow complexity measurement of processes and Weyuker's properties. In *Proceedings of the 6th International Enformatika Conference*, Vol. 8, pp. 213-218, 2005.
20. Krogstie, J. Quality of business process models. In *Proceedings of the IFIP Working Conference on The Practice of Enterprise Modeling*, pp. 76-90, Springer, 2012.
21. Pesic, M., Schonenberg, H., Van der Aalst, W. M. Declare: Full support for loosely-structured processes. In *Proceedings of the 11th IEEE International Enterprise Distributed Object Computing Conference*, EDOC, pp. 117-120, IEEE, 2007.

Author Biographies

Nesrine Missaoui Doctor in computer science and researcher at RIADI laboratory in University of Mannouba, Tunisia. She obtained her thesis degree in Computer Science, in March 2022, at the Higher Institute of Computer Science and Communication Technologies of Hammam Sousse, University of Sousse. Her thesis work focuses on process improvement based on a business process improvement patterns. The author is specialized in the fields of Business Process Management, Process improvement, Process mining, Process simulation, Process sustainability, Pattern application, and e-health.

Maissa Rhif Received her Msc in information and decision support system from the Higher Institute of Management of Souse, University of Sousse, Tunisia. Her research focuses on process discovery and modelling by analyzing the behavior of a process to identify its type and propose ideas for improvement. Her research interests include Business Process Management, Process discovery, Process modelling and Process improvement.

Sonia Ayachi Ghannouchi Full Professor in Business Computing at the Higher Institute of Management of Sousse, in University of Sousse, Tunisia. She obtained her PhD in Computer Science from the University of Mannouba in Tunisia and her HDR in Enterprise Computing from the University of Sousse in Tunisia. Her research interests include: Software engineering and Reengineering, Business Process Modelling, Business Process Management, Process Mining, e-learning and e-health.