Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 225 (2023) 2067–2075

Procedia Computer Science

www.elsevier.com/locate/procedia

27th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2023)

Applying the SIM Tool in Clinical Practice: a Case Study in Neonatal Resuscitation Simulation[1]

Alessio Bottrighi [a], Marco Guazzone [b,c], Giorgio Leonardi [a], Stefania Montani [a], Manuel Striani [a,*], Paolo Terenziani [a]

[a]*Laboratorio Integrato di Intelligenza Artificiale e Informatica in Medicina DAIRI; Azienda Ospedaliera SS. Antonio e Biagio e Cesare Arrigo, Alessandria e DISIT - Università del Piemonte Orientale, Alessandria, Italy*

[b]*Consorzio Nazionale Interuniversitario per le Telecomunicazioni (CNIT), Parma, Italy*

[c]*Computer Science Institute, DISIT, Università del Piemonte Orientale, Alessandria, Italy*

{*alessio.bottrighi, marco.guazzone, giorgio.leonardi, stefania.montani, manuel.striani, paolo.terenziani*}*@uniupo.it*

Abstract

In medical process mining, specific domain characteristics have to be dealt with: in particular, in medicine, a significant amount of expert knowledge is typically available; moreover, an interactive approach, letting medical users be involved in the work of process model discovery, is more acceptable than a completely automated strategy.

To this end, in our recent work we have defined SIM (Semantic Interactive Miner), an innovative process mining tool able to: (i) support the interaction with medical experts, who can progressively merge parts of the initially mined model, obtaining a more generalized version; (ii) exploit pre-encoded domain knowledge, to move from a model where activities are reported at the ground level to a more user-interpretable high-level version.

In this paper we illustrate the features of our tool by showing its application to the case study of neonatal resuscitation simulation: we use SIM to mine the process models produced by two different groups of students of a simulation course, aiming at verifying whether differently skilled young professionals produce different processes, which can finally be compared to the correct guideline.

Peer-review under responsibility of the scientific committee of the 27th International Conference on Knowledge Based and Intelligent Information and Engineering Systems

Keywords: Medical Process Mining, Semantic abstraction, Interactivity;

---

[1] This research has financial support of the University of Piemonte Orientale.
* Corresponding author. Tel.: +39 0131360179
E-mail address: manuel.striani@uniupo.it

1877-0509 © 2023 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer-review under responsibility of the scientific committee of the 27th International Conference on Knowledge Based and Intelligent Information and Engineering Systems
10.1016/j.procs.2023.10.197

1. Introduction

The Process Mining Manifesto [1] highlights that process model discovery is one of the main tasks in process mining: it takes as input an event log, storing the process traces, i.e., the sequences of activities generated during the execution of single process instances, and outputs a process model, which should be obtained "without using any a-priori information."

While such a definition is applicable to several business processes domains, it looks quite restrictive in the medical field, where a lot of domain knowledge is available, both in the form of analysts' knowledge and expertise, and in the form of pre-encoded ontological knowledge (e.g., ATC, Snomed). Moreover, in medicine, an interactive approach, that allows medical experts to be involved in the work of process model discovery, seems more acceptable than a completely automated strategy.

To take into account such needs, and to take advantage of all the available knowledge, in our recent work we proposed the *Semantic Interactive Miner* (SIM) [4, 5], an innovative process mining tool able to:

- support the interaction with medical experts, who can selectively and progressively merge parts of an initial automatically mined model, obtaining a more generalized and reduced version;
- exploit pre-encoded knowledge to move from a model where activities are reported at the ground level to a more user-interpretable and abstract version.

In this paper, we illustrate the characteristics of SIM, by showing its application to the case study of neonatal resuscitation simulation. Specifically, we adopt SIM to mine the process models produced by the traces generated by two different groups of students during a simulation course. We then compare the two process models, to verify whether differently skilled young professionals produce different results, that may have a different level of compliance with the correct guideline.

The paper is organized as follows: in Section 2 we provide a technical description of the main SIM features; in Section 3 we illustrate the case study; in Section 4 we summarize some related work; finally, Section 5 is devoted to conclusions.

2. The SIM tool

In this section, we overview the main features of SIM. For more details, please see our previous works [4, 5].

Most process discovery systems in the literature act as "black-boxes" that, on the basis of a set of parameters, take as input an event log and provide as output a model. If analysts are not satisfied with the model, they can only run the system again (and again), with different parameters. However, since the effects of parameter values on the output process model are totally unknown to the analysts, this iteration with the system looks like a "blind search." On the other hand, SIM aims to provide the opportunity to interactively exploit the analyst's knowledge in model discovery and to consider also pre-encoded domain knowledge, when available.

Specifically, as regard domain knowledge, SIM considers taxonomic knowledge about:

- the refinement (*ISA* relations) of process activities, and
- the composition (*Part-Of* relations) of process activities.

The core idea of SIM is to automatically mine (from the event log) an initial process model, and then incrementally and interactively (with the analysts) generalize it, by applying a set of operations, possibly in different steps. Four main aspects characterize SIM's definition and behavior:

1. *Initial model*. SIM starts by mining from the log an initial model with precision = 1 and replay-fitness = 1 [6]. Informally speaking, it is a model that exactly covers all and only the behavior described by the input traces in the log. Of course, since logs are incomplete, some forms of generalizations are needed. However, considering that generalizations may correspond to a loss of precision and/or replay-fitness, we want that generalizations are directly driven by the analyst, through an interactive and incremental process.

2. *Generalizations*. In SIM, generalizations can be obtained by applying two different types of operations:

   - *merge operations*, which merge the instances of some paths occurring in the model. Notably, to support merge operations, SIM provides a set of facilities to specify paths, to search their occurrences in the model, and to merge them;
   - *abstraction operations*, which move from a process description where activities are reported at the ground level, to more user-interpretable/compact descriptions, in which sets of such activities are abstracted into the "macro-activities" subsuming them (ISA relations) or constituted by them (Part-Of relations).

3. *Incremental and selective application of generalization operators*. SIM provides analysts with the possibility of:

   - applying generalization in a *selective way* (e.g., the analyst can decide what occurrences of the path to merge, or what instances of the actions composing a macro-action to abstract);
   - operating *step-by-step*; in such a way, the final model can be built by applying subsequent merge and abstraction steps, each one leading to a progressively more generalized version.

4. *Model evaluation and versioning*. SIM enables analysts to:

   - perform a quantitative (considering parameters like replay-fitness, precision, generalization [6]) *evaluation* of the model;
   - backtrack to previous versions of the model (*versioning*).

To achieve the above design goals, the architecture of SIM is comprised of five main modules: (1) the *mining module*, which constructs an initial process model from traces in the event log; (2) the *retrieval module*, which looks for paths in the (initial or partially generalized) process model; (3) the *merge module*, which merges (potentially, in a selective way) paths found in a process model by applying one or more *merge operations*; (4) the *abstraction module*, which abstracts (either automatically or interactively) a set of activities of a process model into macro-activities by applying one or more *abstraction operations*, and (5) the *versioning module*, which creates and maintains a tree-like data structure (the *versioning tree*), storing the evolution of a process model, from the initial model to the current one, so that the analyst can further refine a particular version of the model as needed.

3. A case study in neonatal resuscitation simulation

The case study we are reporting in this paper stems from our collaboration with the pediatricians of the neonatal intensive care unit of the Casale Monferrato Hospital in Italy. The hospital is organizing simulation courses to train younger professionals in neonatal resuscitation. The simulation courses exploit Laerdal "Newborn Anne©" device, which has been integrated with ad-hoc software to record all the activities executed during the resuscitation procedures, thus generating an event log. This event log is currently composed of a total of 50 traces. The expert pediatricians provided us with an ontology of the activities to be performed, describing the composition (Part-Of) or subsumption (ISA) relations that define the higher-level procedures.

We have used SIM to mine from the logs the process models produced by two different groups of students of the courses, aiming at verifying whether the differently skilled young professionals produce different resuscitation processes, to compare them with the ideal model (i.e., the clinical guideline), which was available.

Fig. 1 shows the ontology. The main "top-level" procedures to be executed are an evaluation of the baby's situation (*InitialSteps*), a macro-activity (*StartVM*) consisting of ventilation, cardiac support and monitoring, and respiratory support (*SupportRP*). The three composite procedures we mentioned are progressively decomposed through Part-Of relations down to the ground activities. The ideal model (from the guideline) describes the process as follows: *InitialSteps* is the first macro-activity, which has to be followed by *StartVM*; if needed, *SupportRP* can also be performed.

Fig. 2 shows the process model mined by SIM on the 25 traces of the group of the most skilled students, where ground activities (level 0 in the ontology) have already been abstracted to the first level of the ontology. Since several common paths can be easily recognized, we have proceeded with a merge step, obtaining the process model in Fig. 3.

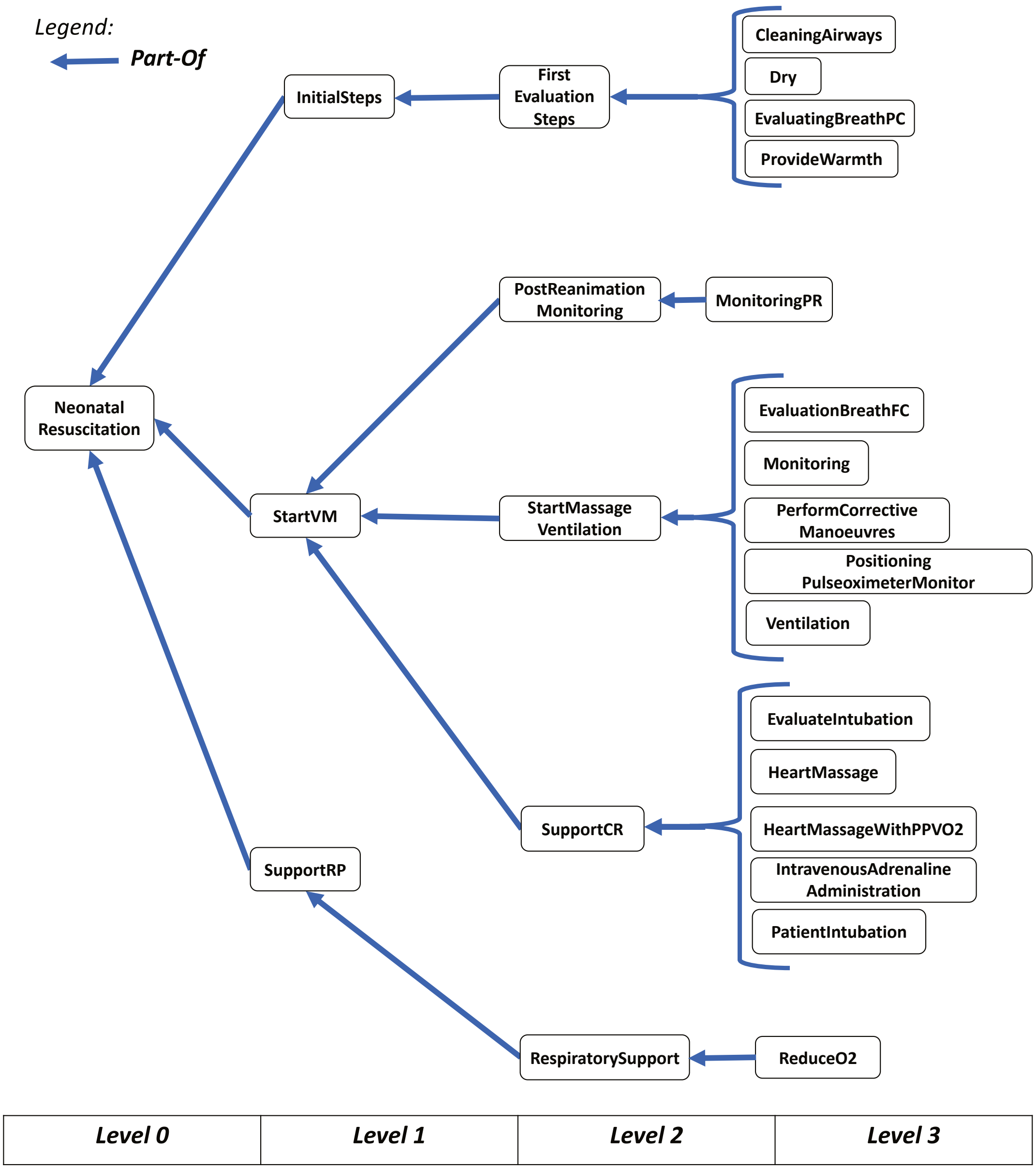

Fig. 1. The knowledge base describing the main neonatal resuscitation management phases

In Fig. 3, we have manually highlighted the items in Part-Of relation that can be further abstracted to the second level of the ontology. As the figure shows, this student group successfully implemented a resuscitation strategy far from the ideal one, which should involve just the sequence "*InitialSteps* → *StartVM*." The lack of expertise has led the group to repeat some procedures and to introduce a couple of additional activities, which cannot be abstracted. However, globally speaking, the teacher has judged their work as a good one.

Fig. 4, on the other hand, shows the process model mined by SIM on the 25 traces of the less skilled students, while Fig. 5 shows the result of the merge procedure, as before. In this case, the steps of the ideal process cannot be easily recognized, demonstrating the significant lack of skill of this student group, which repeats the same macro-activities several times (e.g., *SupportCR* is executed before and after *StartMassageVentilation*, while it should be executed only after it), and needs to introduce respiratory support with oxygen (*RespiratorySupport*) – because the baby simulator, incorrectly cared, reveals unstable clinical conditions.

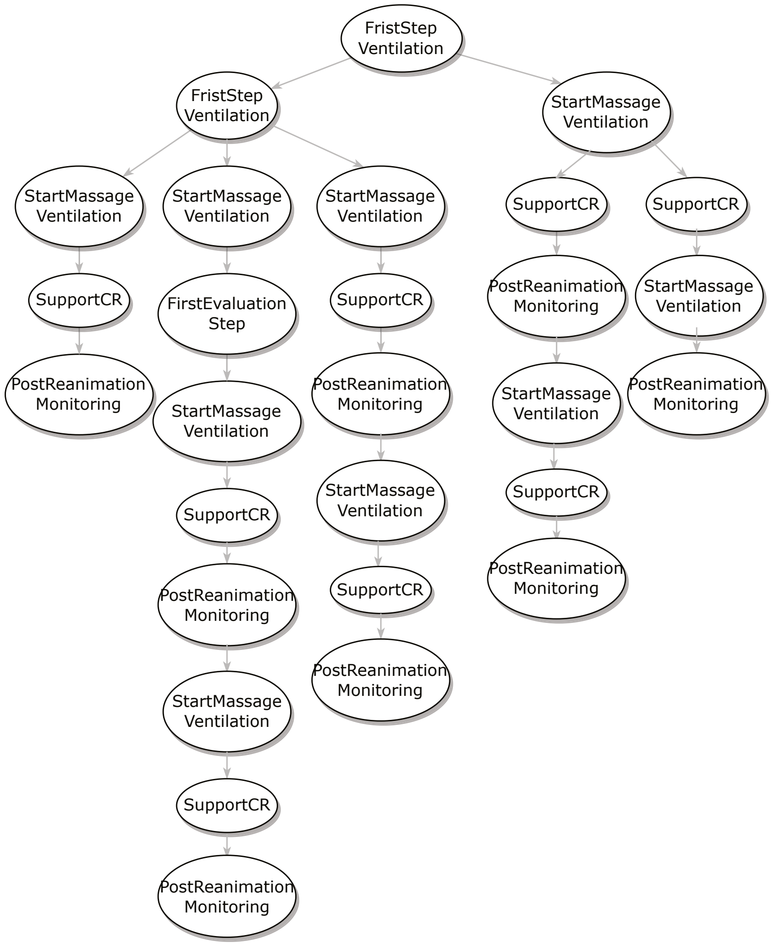

Fig. 2. Process model of the most skilled students, abstracted at the first level of the ontology. The process starts from an initial activity (FirstStepVentilation) and splits have to be interpreted as mutually exclusive

This experimental work shows an important feature of our abstraction functionality: while simplifying process readability, providing a bird's eye view of the procedures, it is still able to highlight important differences among the compared models. Notably, in our experimental work, less relevant details are lost, but key pieces of information are kept.

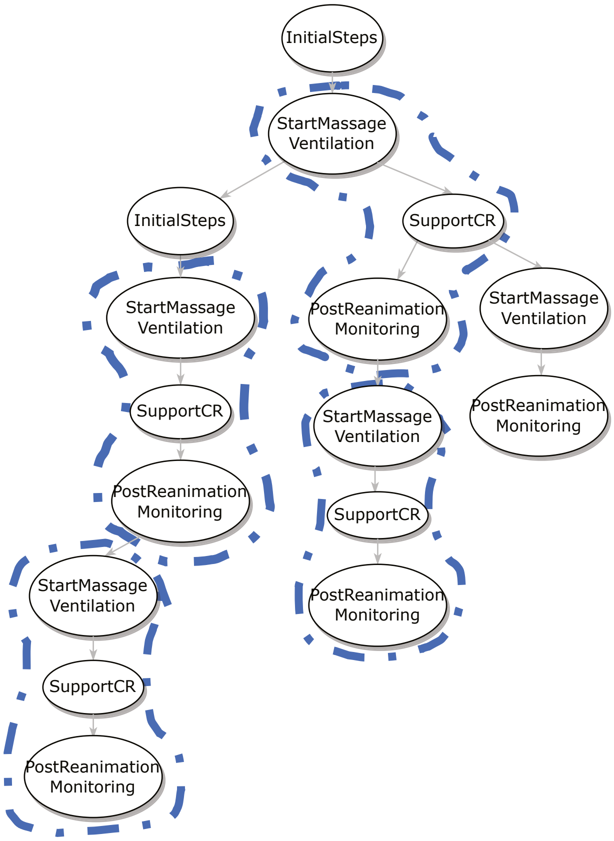

Fig. 3. Process model of the most skilled students, where further possible abstractions at the second level of the ontology are highlighted. For instance, activities {*StartMassageVentilation*, *SupportCR*, *PostReanimationMonitoring*} are Part-Of **StartVM**

4. Related work

Various recent contributions consider how to integrate semantic information and analyst's expertise into process discovery. In particular, the work in [2] proposes a two-step approach, where, first, one or more "low-level models" are constructed, and then the selected one is converted into a "high-level model", able to express more advanced control flow patterns. In the first step, several syntactic and semantic abstraction techniques are proposed; in the second step, an analyst is allowed to evaluate the models, and use her/his knowledge to judge which one guarantees a sufficient level of generalization. SIM moves further in the direction started by this work, giving additional power and role to the human expert, and incorporating ontological knowledge.

The work in [10], on the other hand, identifies groups of related activities, corresponding to abstract tasks. The abstraction mechanism exploits a clustering technique, which adopts activity attributes (e.g., cost, or resources) to cluster the group of ground activities into the corresponding abstract task. The approach requires all activity attributes to be logged, but unfortunately this information is often incomplete in practice, especially in the medical field.

The work in [8] proposes an approach which encodes domain knowledge as a set of precedence constraints. The work in [9] focuses on declarative process models and prunes constraints that, based on domain knowledge, can be inferred from other constraints. It is however worth noting that, in all of these contributions, domain knowledge must necessarily be expressed in the form of rules or constraints, and process discovery can be very complex.

A very interesting contribution is described in [7], which introduces Interactive Process Discovery (IPD). IPD exploits domain knowledge and the event log in parallel, improving accuracy and understandability. To enable interactive discovery, IPD uses the so-called synthesis rules, allowing to expand a minimal synthesized net (a type of Petri Net) by adding one transition and/or one place, selected by the user, at a time. IPD indicates to the user if the selected

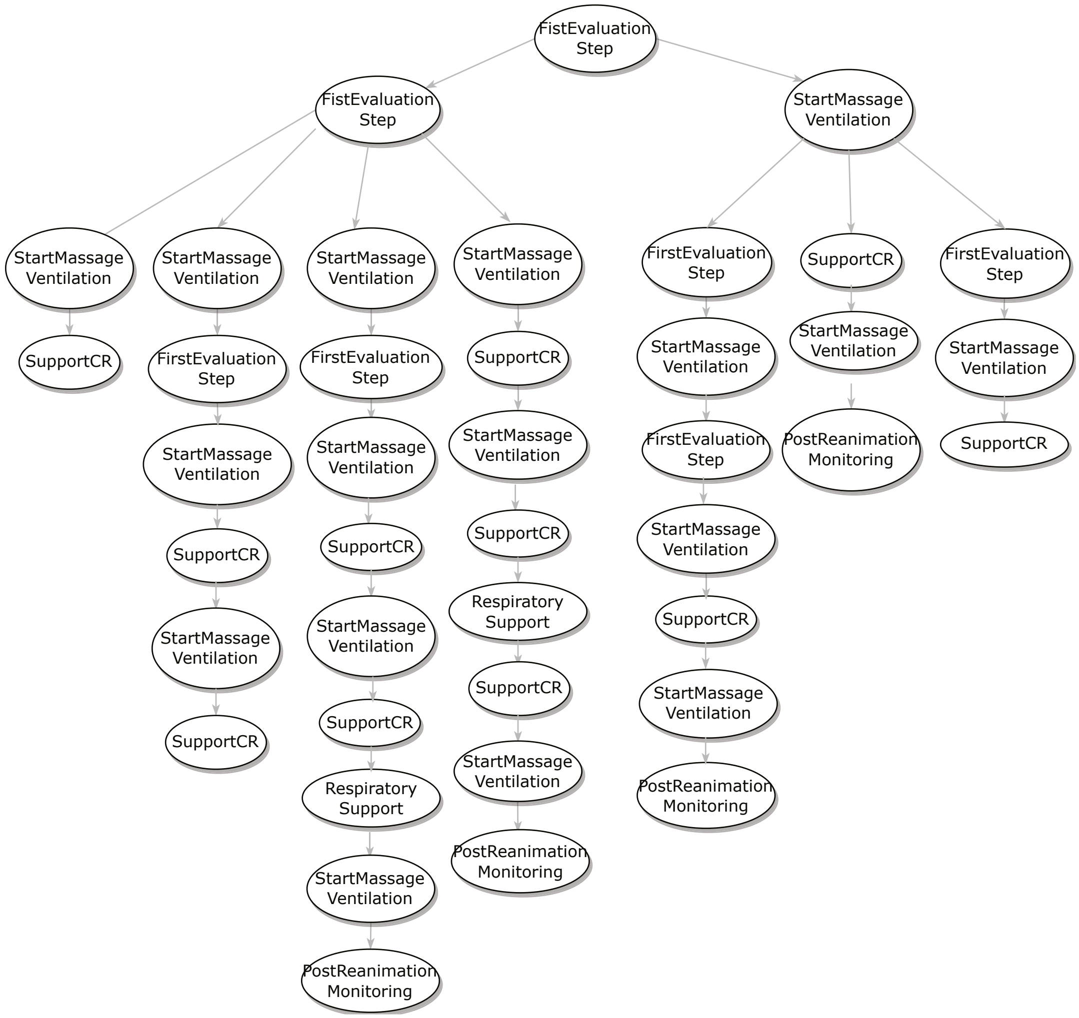

Fig. 4. Process model of the less skilled students, abstracted at the first level of the ontology.

activity occurs before or after the other activities within the synthesized net, and suggests where to place it, based on the log. The paper in [3] discussed the suitability of IPD to healthcare. The goals of such an approach are probably the closest to SIM, even if it adopts a different technical solution.

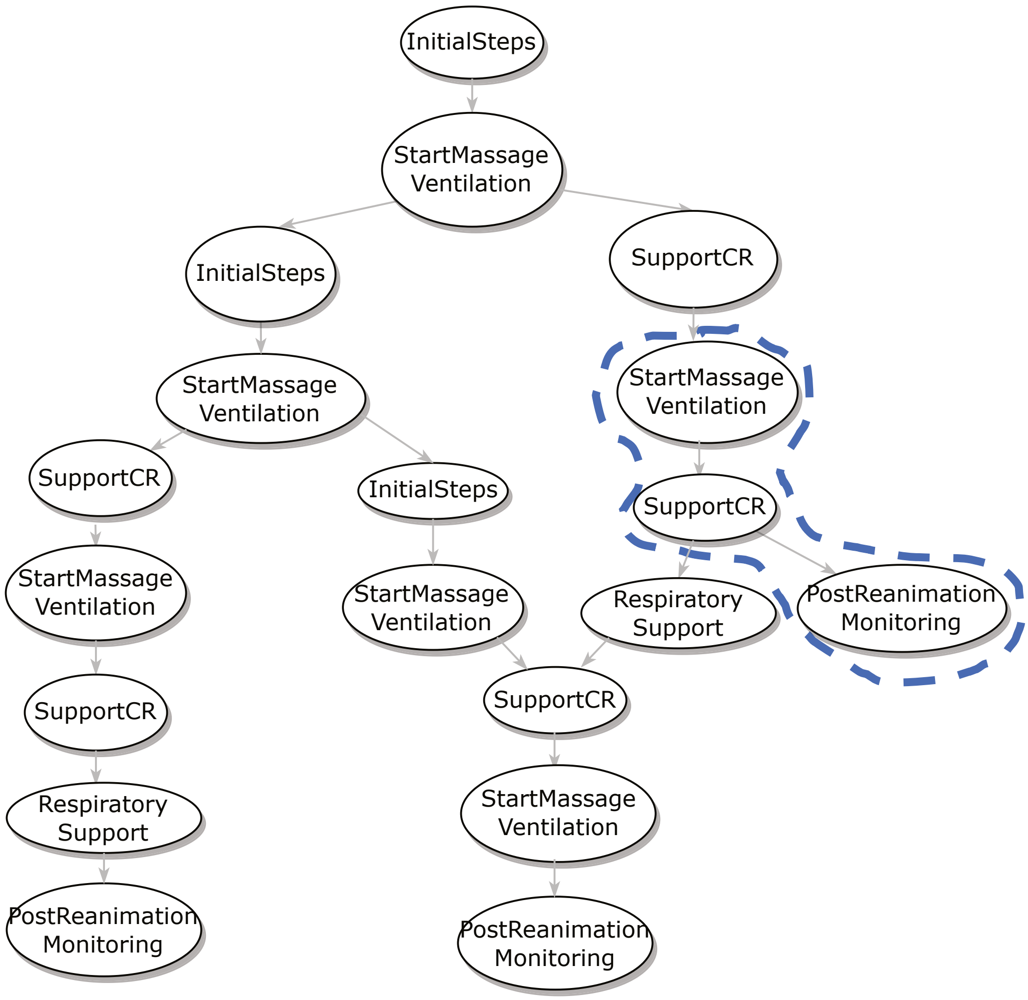

Fig. 5. Process model of the less skilled students, where further possible abstractions at the second level of the ontology are highlighted.

5. Conclusions

In this paper, we have illustrated the behaviour of SIM, a process discovery tool specifically designed for medical applications. Indeed, SIM allows to deal with the significant amount of knowledge typically available in medicine, both incorporating it in the ontological form, and by allowing interaction with medical analysts, thus supporting progressive process model generalization.

Our case study in neonatal resuscitation simulation has shown the potentialities of SIM: on the one hand, the tool increases process readability and understandability, thanks to its merging and abstraction operations, which remove less relevant details; on the other hand, it is able to highlight important differences among the compared models, by keeping all the more relevant information.

In the future, we plan to evaluate SIM in different medical domains as well, considering usability issues as well.

Acknowledgments

This research has been partially supported by the INDAM – GNCS Project 2023.

References

[1] van der Aalst, W., Adriansyah, A., de Medeiros, A.K.A., Arcieri, F., Baier, T., Blickle, T., Bose, J.C., van den Brand, P., Brandtjen, R., Buijs, J., Burattin, A., Carmona, J., Castellanos, M., Claes, J., Cook, J., Costantini, N., Curbera, F., Damiani, E., de Leoni, M., Delias, P., van Dongen, B.F., Dumas, M., Dustdar, S., Fahland, D., Ferreira, D.R., Gaaloul, W., van Geffen, F., Goel, S., Günther, C., Guzzo, A., Harmon, P., ter Hofstede, A., Hoogland, J., Ingvaldsen, J.E., Kato, K., Kuhn, R., Kumar, A., La Rosa, M., Maggi, F., Malerba, D., Mans, R.S., Manuel, A., McCreesh, M., Mello, P., Mendling, J., Montali, M., Motahari-Nezhad, H.R., zur Muehlen, M., Munoz-Gama, J., Pontieri, L., Ribeiro, J., Rozinat, A., Seguel Pérez, H., Seguel Pérez, R., Sepúlveda, M., Sinur, J., Soffer, P., Song, M., Sperduti, A., Stilo, G., Stoel, C., Swenson, K., Talamo, M., Tan, W., Turner, C., Vanthienen, J., Varvaressos, G., Verbeek, E., Verdonk, M., Vigo, R., Wang, J., Weber, B., Weidlich, M., Weijters, T., Wen, L., Westergaard, M., Wynn, M., 2012. Process mining manifesto, in: Daniel, F., Barkaoui, K., Dustdar, S. (Eds.), Business Process Management Workshops, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 169–194. doi:10.1007/978-3-642-28108-2_19.

[2] van der Aalst, W.M.P., Rubin, V.A., Verbeek, H.M.W., van Dongen, B.F., Kindler, E., Günther, C.W., 2010. Process mining: a two-step approach to balance between underfitting and overfitting. Softw. Syst. Model. 9, 87–111. doi:10.1007/s10270-008-0106-z.

[3] Benevento, E., Dixit, P.M., Sani, M.F., Aloini, D., van der Aalst, W.M.P., 2019. Evaluating the effectiveness of interactive process discovery in healthcare: A case study, in: Francescomarino, C.D., Dijkman, R.M., Zdun, U. (Eds.), Business Process Management Workshops - BPM 2019 International Workshops, Vienna, Austria, September 1-6, 2019, Revised Selected Papers, Springer. pp. 508–519. doi:10.1007/978-3-030-37453-2_41.

[4] Bottrighi, A., Canensi, L., Leonardi, G., Montani, S., Terenziani, P., 2018. Interactive mining and retrieval from process traces. Expert Syst. Appl. 110, 62–79. doi:10.1016/j.eswa.2018.05.041.

[5] Bottrighi, A., Guazzone, M., Leonardi, G., Montani, S., Striani, M., Terenziani, P., 2022. Integrating isa and part-of domain knowledge into process model discovery. Future Internet 14. doi:10.3390/fi14120357.

[6] Buijs, J.C.A.M., van Dongen, B.F., van der Aalst, W.M.P., 2012. On the role of fitness, precision, generalization and simplicity in process discovery, in: Meersman, R., Panetto, H., Dillon, T.S., Rinderle-Ma, S., Dadam, P., Zhou, X., Pearson, S., Ferscha, A., Bergamaschi, S., Cruz, I.F. (Eds.), On the Move to Meaningful Internet Systems: OTM 2012, Confederated International Conferences: CoopIS, DOA-SVI, and ODBASE 2012, Rome, Italy, September 10-14, 2012. Proceedings, Part I, Springer. pp. 305–322. doi:10.1007/978-3-642-33606-5_19.

[7] Dixit, P.M., Verbeek, H.M.W., Buijs, J.C.A.M., van der Aalst, W.M.P., 2018. Interactive data-driven process model construction, in: Trujillo, J., Davis, K.C., Du, X., Li, Z., Ling, T.W., Li, G., Lee, M. (Eds.), Conceptual Modeling - 37th International Conference, ER 2018, Xi'an, China, October 22-25, 2018, Proceedings, Springer. pp. 251–265. doi:10.1007/978-3-030-00847-5_19.

[8] Greco, G., Guzzo, A., Lupia, F., Pontieri, L., 2015. Process discovery under precedence constraints. ACM Trans. Knowl. Discov. Data 9, 32:1–32:39. doi:10.1145/2710020.

[9] Maggi, F.M., Bose, R.P.J.C., van der Aalst, W.M.P., 2013. A knowledge-based integrated approach for discovering and repairing declare maps, in: Salinesi, C., Norrie, M.C., Pastor, O. (Eds.), Advanced Information Systems Engineering - 25th International Conference, CAiSE 2013, Valencia, Spain, June 17-21, 2013. Proceedings, Springer. pp. 433–448. doi:10.1007/978-3-642-38709-8_28.

[10] Smirnov, S., Reijers, H.A., Weske, M., 2012. From fine-grained to abstract process models: A semantic approach. Inf. Syst. 37, 784–797. doi:10.1016/j.is.2012.05.007.