

# Weighted violations in alignment-based conformance checking<sup>\*</sup>

Joscha Gröger<sup>1,2</sup>[0000-0001-7538-1248], Tobias Geyer<sup>2</sup>[0000-0003-3072-1709],  
Martin Kuhn<sup>2</sup>[0000-0002-3242-1251], Stephan A. Braun<sup>3,4</sup>[0000-0002-5600-0195],  
and Ralph Bergmann<sup>1,2</sup>[0000-0002-5515-7158]

<sup>1</sup> Business Information Systems II, University of Trier, Trier, Germany

<sup>2</sup> German Research Center for Artificial Intelligence (DFKI), Branch Trier, Germany

<sup>3</sup> Dept. of Dermatology, University Hospital Münster, Germany

<sup>4</sup> Dept. of Dermatology, Heinrich Heine University, Düsseldorf, Germany

**Abstract.** Conformance checking is a process mining technique that allows verifying the conformance of process instances to a given model. Many conformance checking algorithms provide quantitative information about the conformance of a process instance through metrics such as fitness. Fitness measures to what degree the model allows the behavior observed in the event log. Conventional fitness does not consider the individual severity of deviations. In cases where there are rules that are more important to comply with than others, fitness consequently does not take all factors into account. In the field of medicine, for example, there are guideline recommendations for clinical treatment that have information about their importance and soundness, making it essential to distinguish between them. Therefore, we introduce an alignment-based conformance checking approach that considers the importance of individual specifications and weights violations. The approach is evaluated with real patient data and evidence-based guideline recommendations. Using this approach, it was possible to integrate guideline recommendation metadata into the conformance checking process and to weight violations individually.

**Keywords:** process mining, conformance checking, alignments, fitness, weighted violations, guideline compliance

## 1 Introduction

Process mining is an emerging research field and fills the gap between data mining and business process management [3]. One technique of process mining is conformance checking, whose approaches focus on measuring the conformance of a process instance to a process model. The results of the measurement can usually be output in the form of alignments [2], i.e., corrective adjustments for process instances, or metrics. A common metric is fitness, which measures

---

<sup>\*</sup> The research is funded by the German Federal Ministry of Education and Research (BMBF) in the project DaTreFo under the funding code 16KIS1644.

to what degree the model allows the behavior observed in the event log that contains all process instances.

However, conventional fitness values each deviation, i.e., a rule violation against the specified behavior, equally. This becomes problematic in terms of assessment for use cases in which there are rules that are more important than others and, consequently, a violation of them is also worse. For instance, in the domain of medicine, there are clinical guidelines. Clinical guidelines are systematically developed statements that reflect the current state of medical knowledge to support physicians and patients in the decision-making process for appropriate medical care in specific clinical situations [11]. These statements have metadata (e.g., level of evidence or consensus strength) that provide information about their importance and soundness. Therefore, it is important to distinguish between the degree of deviation and to weight rule violations differently in order to obtain more accurate and meaningful results. In a scoping review, Oliart et al. [13] systematically assessed the criteria used to measure adherence to clinical guidelines and examined the suitability of process mining techniques. So far, there is no approach that allows different weighting of guideline statements [13]. Therefore, in this paper, we present a first approach for weighted violations in alignment-based conformance checking that incorporates the assessment of individual specifications in the calculation of fitness.

The approach is a promising solution to address medicine-specific characteristics and challenges for process mining presented in Munoz-Gama et al. [12]. Regarding the characteristics, we deal with the use of guidelines (D3) in the process mining context. In particular, concrete characteristics of guidelines are integrated to generate more valuable results. Furthermore, we built on characteristic D5, the consideration of data at multiple abstraction levels, by also integrating medical metadata. In addition, our research involves healthcare professionals (D6) who have made a valuable contribution to its realization. Regarding challenges, we address dealing with reality (C4) as we test and evaluate our approach with real patient data. Furthermore, the development of this approach should foster the use of process mining by healthcare professionals (C5), as it leads to helpful and valuable results.

The remainder of the paper is organized as follows. Section 2 provides background information on the components of our approach. Section 3 describes the methodological approach for the alignment-based conformance checking with weighted violations. Section 4 presents the evaluation process. In Section 5, the findings are discussed, and Section 6 concludes the paper.

## 2 Fundamentals

### 2.1 Event Logs

Process mining is based on event logs. Event logs can be viewed as multi-sets of cases. Each case consists of a sequence of events, i.e., the trace. Events are execution instances of activities. Here, the execution of an activity can be represented by multiple events. This can occur, for example, when multiple lifecycle

stages of execution are logged [15]. In addition to the control flow perspective, event logs can also use attributes to represent other perspectives, such as the data perspective or the resource perspective. The following defines event logs, traces, events, attributes, and functions on them as a basis for the methodology.

**Definition 1.** (*Universes*) For this paper, we define the following universes:

- $\mathcal{V}$  is the universe of all possible variable identifiers
- $\mathcal{C}$  is the universe of all possible case identifiers
- $\mathcal{E}$  is the universe of all possible event identifiers
- $\mathcal{A}$  is the universe of all possible activity identifiers
- $\mathcal{AN}$  is the universe of all possible attribute identifiers

**Definition 2.** (*Attributes, Classifier*) Attributes can be used to characterize events and cases, e.g., an event can be assigned to a resource or have a timestamp. For any event  $e \in \mathcal{E}$ , any case  $c \in \mathcal{C}$  and name  $n \in \mathcal{AN}$ ,  $\#_n(e)$  is the value of attribute  $n$  for event  $e$  and  $\#_n(c)$  is the value of attribute  $n$  for case  $c$ .  $\#_n(e) = \perp$  if event  $e$  has no attribute  $n$  and  $\#_n(c) = \perp$  if case  $c$  has no attribute  $n$ . We assume the classifier  $\underline{e} = \#_{\text{activity}}(e)$  as the default classifier.

**Definition 3.** (*Trace, Case*) Each case  $c \in \mathcal{C}$  has a mandatory attribute trace, with  $\hat{c} = \#_{\text{trace}}(c) \in \mathcal{E}^* \setminus \{\langle \rangle\}$ . A trace is a finite sequence of events  $\sigma \in \Sigma^*$  where each event occurs only once, i.e.  $1 \leq i < j \leq |\sigma| : \sigma(i) \neq \sigma(j)$ . By  $\sigma \oplus e = \sigma$  we denote the addition of an  $e$  event to a trace  $\sigma$ .

**Definition 4.** (*Event log*) An event log is a set of cases  $\mathcal{L} \subseteq \mathcal{C}$ , in the form that each event is contained only once in the event log. If an event log contains timestamps these should be ordered in each trace.  $\hat{\mathcal{L}} = \{e | c \in \mathcal{L} \wedge e \in \hat{c}\}$  is the set of all events appearing in the log  $\mathcal{L}$ .

## 2.2 Alignments

To check the conformance of an event log  $\mathcal{L}$  to a process model  $M$ , approaches to search for alignments are common for different process modeling languages [5]. An alignment shows how a log or trace can be replayed in a process model.

**Definition 5 (Alignment, moves).** Let  $\gg$  be the indicator for no move and  $\mathcal{E}_{\gg} = \mathcal{E} \cup \{\gg\}$  the input alphabet including the no move. Then  $\mathcal{E}_A = (\mathcal{E}_{\gg} \times \mathcal{E}_{\gg}) \setminus \{(\gg, \gg)\}$  is the set of legal moves. Let  $(s', s'')$  be a pair of values with  $(s', s'') \in \mathcal{E}_A$ , then holds:

- is a log move if  $s' \in \mathcal{E}$  and  $s'' = \gg$
- is a model move if  $s'' \in \mathcal{E}$  and  $s' = \gg$
- is a synchronous move if  $(s', s'') \in (\mathcal{E} \times \mathcal{E}) \wedge s' = s''$

An alignment of two traces  $\sigma', \sigma'' \in \mathcal{E}^*$  is a sequence  $\gamma \in \mathcal{E}_A^*$ .

In other approaches, the alignment definition may differ from the above. However, the described approach can be adapted for all cost-based alignments.

### 2.3 MLMs and Arden Syntax

Medical Logic Modules are designed to represent medical knowledge in self-contained units that are both human-readable and computer-interpretable. Moreover, they should be transferable between several clinics [4, 10]. The Arden syntax for MLMs allows the development of MLMs. It is a rule-based, declarative, HL7 standardized approach to open implementation of MLMs [14]. This was developed specifically to formalize and exchange medical knowledge. In the following, we interpret the term MLM as MLM in the Arden syntax. MLMs are text files divided into discrete slots (see Figure 1). These slots then contain data, describe database queries or rules [10]. The basic orientation of MLMs are to formalize medical knowledge and to formulate rules, which are usually of the form “If patient has fever  $\geq 40$ , then make a request for examination Z”. This logic is formulated in the so-called logic slot and allows complex queries [4]. Among the operators are also operators with procedural reference like **before**, **after**, **within same day** or **n days before/after**. However, these do not directly compare events, only timestamps.

This approach was repurposed in the paper [8] to check the conformance of treatment sequences. For this purpose, part of the guideline for the treatment of malignant melanoma already used in [9] was transformed into MLMs using the CGK4PM framework [7]. The framework is inspired by the guideline creation process and enables the systematic transformation of guideline knowledge in an iterative procedure involving domain experts. Instead of MLMs being used to establish if-then rules, they were used in the approach to verify whether the particular guideline statement was followed. In case of non-compliance, manually modeled alignment steps were returned, which were then implemented by the client. This approach is used below to evaluate the approach in this paper.

### 2.4 MLM-based conformance checking

To describe our approach, we introduce a simplified formalization of MLMs and the MLM-based conformance checking approach proposed by Grüger et al. [8].

**Definition 6 (MLM and Slot).** *We define an MLM  $m$  as a quadruple consisting of four categories with  $m = (\text{maintenance}, \text{library}, \text{maintenance}, \text{resources})$ . Each category  $c$  consists of predefined slots. Let  $S$  be the set of all slots, then  $S_c \subset S$  is the set of all slots defined for category  $c$ . Each slot consists of one to many values. So  $m[s]$  returns the values of slot  $s$  for MLM  $m$ .*

Each MLM defines in the evoke slot at which evocation event it is evaluated. Here, the term evocation event extends the event concept to include the event classifier and data-level writing events. At the data level, events can be defined by the attribute name or the name in combination with the attribute value.

**Definition 7 (evoke, evocation event).** *Let  $e \in \mathcal{E}$  be an event and  $m$  be an MLM.  $E_e$  defines the set of all evocation events evoked for event  $e$ :*

$$E_e = \{\underline{e}\} \cup \{e_n | n \in \mathcal{AN} \text{ if } e_n \neq \perp\} \cup \{(e_n, \#e_n) | n \in \mathcal{AN} \text{ if } e_n \neq \perp\}$$

Then there exists a function  $evoked_m : \mathcal{E} \cup \mathcal{AN} \cup (\mathcal{AN} \times \mathcal{V}) \rightarrow \{0, 1\}$  with:

$$evoked_m(E_e) = \begin{cases} 1 & \text{if } \#(m[evoked] \cap (x)) > 0 \\ 0 & \text{else} \end{cases}$$

Let  $M \subseteq MLM$  be a declarative model consisting of many MLMs, and  $\sigma$  be a trace. Then holds:

$$M_\sigma = \{m \in M \mid \exists e \in (E) : evoked_m(E_e) = 1\}$$

is the shorthand for all evoked MLMs from  $M$  for  $\sigma$ .

The logic slot defines the actual conformance check based on the trace data from the data slot. The actual logic of the conformance checking and the alignment is adapted from [8] and described as a black box due to lack of space.

**Definition 8 (Logic, Return).** Let  $A$  be the universe of alignment steps,  $K$  the universe of keys used in the slots and  $V$  the universe of values. Let  $m$  be an MLM, then there exists a logical function  $l : MLM \rightarrow \{0, 1\} \times A^* \times (K, V)^*$ . The boolean value specifies whether the MLM was validated to be conform (1) or not (0). The alignment steps describe the steps for aligning a given trace.

**Definition 9 (Fitness).** Let  $MLM$  be the universe of all MLMs,  $M \subseteq MLM$  be the declarative model, and  $\sigma$  be a trace. The function  $eval : M \times \Sigma \rightarrow \{0, 1\}$  evaluates whether a trace conforms to an MLM or not or was not evoked. The fitness is defined as:

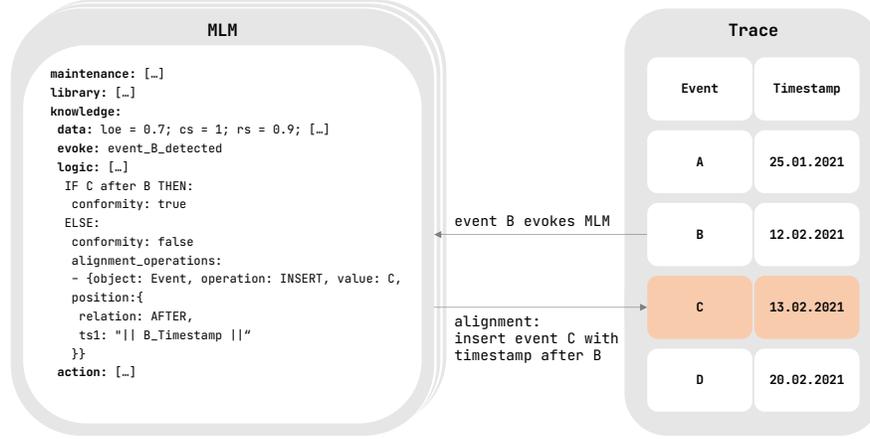
$$fitness(\sigma, M) = \frac{\sum_{i=1}^n eval(\sigma, M_i)}{|M'_\sigma|}$$

An outlined example of an alignment computed with an MLM is shown in Figure 1. Here, event C is supposed to occur after event B. Since event B occurs in the trace, the MLM is evoked. The logic slot concludes to **false** since event C does not occur after event B. Hence, the defined alignment operation in the **else** block is executed and event C is inserted after event B. The timestamps of the events are used to find the correct position for the insertion.

### 3 Methodology

In order to incorporate the degree of a deviation into fitness to consider the importance of the violated part of the model, we introduce an approach to weight the cost of a deviation based on the given metadata. Consequently, we introduce a cost function  $\mathcal{K} : \mathcal{E}_A \rightarrow \mathbb{R}_0^+$ . Here, any cost function can be used that best represents the costs of the particular process and the domain-specific context.

For computing, the fitness of a trace  $\sigma \in \mathcal{L} \subseteq \mathcal{E}^*$  to a process model  $M$  based on the cost function, a complete alignment with minimum cost  $\gamma^{opt}$  is sought. Moreover, the reference alignment  $\gamma_{\sigma_{\mathcal{L}}}^{ref}$  is searched. Thereby, the type of process



**Fig. 1.** Example showing a trace violating the MLM, which states that event B must be followed by C. The alignment step modeled manually in the MLM indicate that C is to be inserted after B.

modeling and the algorithm for calculating the alignment can be individually selected. Typically, the reference alignment with the highest cost is an alignment in which only moves exist in model and log:

$$\gamma_{\sigma_C}^{ref} = \frac{\mathbf{L} \mid a_1^L \mid \dots \mid a_n^L \mid \gg \mid \gg \mid \gg}{\mathbf{M} \mid \gg \mid \gg \mid \gg \mid a_1^M \mid \dots \mid a_n^M}$$

While an alignment is a sequence  $\gamma$  of pairs  $(s', s'') \in (\mathcal{E}_{\gg} \times \mathcal{E}_{\gg}) \setminus \{(\gg, \gg)\}$ , the cost of  $\gamma$  is the sum of the costs of each pair of alignments:

$$\mathcal{K}(\gamma) = \sum_{(s', s'') \in \gamma} \mathcal{K}((s', s''))$$

This is where the approach comes in. Each pair of an alignment  $(s', s'')$  with  $s' \neq s''$  represents a deviation detected by the conformance checking algorithm using the model  $M$ . Therefore, there is a condition  $c$  in the model that caused this violation. We use condition as a term for modeling elements from imperative and declarative approaches (e.g., guards or rules).

**Definition 10 (Condition, Condition weight).** *Let  $M$  be a model and  $C$  be the set of all conditions. Then  $C_M \subseteq C$  is the set of all conditions of  $M$ . Following functions are defined over  $C$ :*

- $w : C \rightarrow \mathbb{R}_0^+$ , the weighting for condition  $c \in C_M$ . As shorthand we use  $w_c = w(c)$ .
- $m_M : (\mathcal{E}_{\gg} \times \mathcal{E}_{\gg}) \setminus \{(\gg, \gg)\} \rightarrow C$ , a mapping of an alignment pair on the condition, causing the violation.

Mapping the alignment pairs  $(s', s'')$  of an alignment  $\gamma$  to a condition  $c$  allows using  $w_c$  to assign a weight from  $\mathbb{R}_0^+$  to each deviation in  $\gamma$  based on  $c$ .

**Definition 11 (violation-weighted cost function).** Let  $\mathcal{E}_A = (\mathcal{E}_{\gg} \times \mathcal{E}_{\gg}) \setminus \{(\gg, \gg)\}$ , then  $\mathcal{K}_{\mathcal{W}} : \mathcal{E}_A \rightarrow \mathbb{R}_0^+$  is the violation-weighted cost function. If  $(s', s'') \in \mathcal{E}_A$ , then

$$\mathcal{K}_{\mathcal{W}}((s', s'')) = w_{m((s', s''))} * \mathcal{K}(s', s'')$$

calculates the weighted cost for the alignment pair  $(s', s'')$ .

**Definition 12 (violation-weighted fitness function).** Let  $\sigma_L \in \mathcal{E}^*$  be a log trace and  $M$  a model. Let  $\gamma_{\sigma_L}^{opt} \in \mathcal{E}_A^*$  be an optimal alignment of  $\sigma_L$  and model  $M$  and  $\gamma_{\sigma_L}^{ref}$  the reference alignment. The fitness level is defined as follows:

$$\mathcal{F}_{\mathcal{W}}(\sigma_L, M) = 1 - \frac{\mathcal{K}_{\mathcal{W}}(\gamma_{\sigma_L}^{opt})}{\mathcal{K}_{\mathcal{W}}(\gamma_{\sigma_L}^{ref})}$$

Therefore, for each deviation in the optimal alignment  $\gamma_{\sigma_L}^{opt}$  and in the reference alignment  $\gamma_{\sigma_L}^{ref}$ , the deviation weighted cost is calculated. This enables algorithm and process modeling language independent for all alignment-based conformance checking approaches to reflect the importance of violated rules in the fitness level.

## 4 Evaluation

For the evaluation, we used the data and model base from Grüger et al. [8]. In this paper, the authors present an MLM-based approach to conformance checking for clinical guidelines. Clinical guidelines are intended to support evidence-based treatment of patients. As a summary of systematically developed recommendations based on extensive literature studies, they are intended to optimize treatment of patients based on evidence [6]. In the original approach [8], part of the guideline for the treatment of malignant melanoma [1] (diagnosis and therapy in primary care and locoregional metastasis) was modeled as a declarative rule-based MLM model. We use this and the dataset consisting of five real patients from the University Hospital Münster to evaluate the approach described. This ensures immediate comparability with the conformance checking results from the original MLM-based conformance checking approach.

In addition, medical guidelines inherently contain information on the timeliness, importance, and foundation of each guideline recommendation, which could not be addressed in previous conformance checking approaches. Therefore, we adapt the approach to compute the violation-weighted fitness such that the weights are dynamically derived from the properties of the guideline statements represented by the MLMs. For calculation, we use the attributes *level of evidence*, *date of last review*, *consensus strength*, and *recommendation strength*.

- **level of evidence (loe):** evidence grading is according to Oxford (2009 version) and is divided into 10 grades (1a, 1b, 1c, 2a, 2b, 2c, 3a, 3b, 4, 5), with 1a (systematic review with homogeneity of randomized-controlled trials) highest loe and 5 (expert opinion without critical analysis or based on physiologic or experimental research or “basic principles”) lowest loe.

- **consensus strength (cs)**: indicates the strength of consensus in the expert panel on the respective statement in percent.
- **recommendation strength (rs)**: for all recommendations, the strength of the recommendation is expressed as A (strong recommendation), B (recommendation), and C (recommendation open).
- **date of last review (dolr)**: indicates the year of the last review of the statement. Considering constant progress, the topicality of recommendations is to be taken into account in the evaluation.

In order to incorporate the weighting attributes  $WA = \{loe, cs, rs, dolr\}$  as weights into the fitness calculation, the individual classification values are mapped as values between 0 and 1, using the mapping function  $m$ . Let  $C$  be the set of MLMs. Let  $m : C \times WA \rightarrow [0, 1]$  be the mapping function for the weighting attributes for a concrete condition  $c \in C$ . For each of the weighting attributes,  $m$  is defined as follows.

For the 10-step gradation of the level of evidence (loe), the values are descending equally distributed over the range from 0 to 1. The strength of recommendation (rs) can be expressed by three different categorical values. Accordingly, the weighting is given in thirds of steps.

loe	1a	1b	1c	2a	2b	2c	3a	3b	4	5	rs	A	B	C
$\mathbf{m(c,loe)}$	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	$\mathbf{m(c,rs)}$	1	0.66	0.33

Since consensus strength (cs) is expressed in relative values from 0 to 100 percent, the mapping values are divided by 100. The date of the last review is divided into time intervals. Recommendations that have been reviewed since 2019 receive the highest recommendation. Review years below that receive a weight of 0.8. This expresses the strength of differentiating fine-grained between the informative value of the individual attributes. For example, the last review year was rated as less relevant by the domain experts.

$$m(c, cs) = \frac{c_{cs}}{100} \quad m(c, dolr) = \begin{cases} 1 & \text{if } c_{dolr} \geq 2018 \\ 0.8 & \text{else} \end{cases}$$

Furthermore, it is necessary to differentiate between standard and critical MLMs. In a critical MLM,  $loe$ ,  $cs$ ,  $rs$  and  $dolr$  are all equal to 1. This means that this MLM is up-to-date and is seen as critical by medical experts. Thus, it is necessary to increase the weight of these MLMs. This is guaranteed by using the function below. In the case the MLM is critical, the defined if-condition holds and the value of 2 is assigned as weight. If the MLM is not critical, the else-condition is applied and the weight for a given MLM  $c$  is calculated as the sum of the mapped values  $v \in WA$  divided by the number of weighting attributes.

$$w(c) = \begin{cases} 2 & \text{if } \sum_{a \in WA} m(c, a) = |WA| \\ \frac{\sum_{a \in WA} m(c, a)}{|WA|} & \text{else} \end{cases}$$

Grüger et al.’s approach [8] returns a semantically optimal alignment. This is manually pre-modeled for each of the MLMs and addresses violations of the MLMs in such a way that they are correctly resolved from a medical perspective, i.e., no overwriting of values in the data perspective, no changing of the guideline model, no most favorable path (e.g., by deleting nodes). This optimal alignment is then incorporated into the calculation of fitness in the denominator. Since the approach is built based on a set of rules in the form of MLMs, but not all of them are evoked for each trace, the reference alignment is computed based only on the set of evoked MLMs  $M_\sigma$  for the trace  $\sigma$  (see Definition 7).

Therefore, we adapt the fitness function established in Definition 9 and modify it as follows. For trace  $\sigma$  and the MLM-based model  $M$ ,  $\gamma_\sigma^{opt}$  is the optimal alignment. Then  $\gamma_\sigma^{ref}$  is the reference alignment violating every MLM in  $M_\sigma$ .

$$\mathcal{F}_W(\sigma_L, M) = 1 - \frac{\mathcal{K}_W(\gamma_\sigma^{opt})}{\mathcal{K}_W(\gamma_\sigma^{ref})} \quad \mathcal{K}_W((s', s'')) = (w_{c((s', s''))})^2 * \mathcal{K}(s', s'')$$

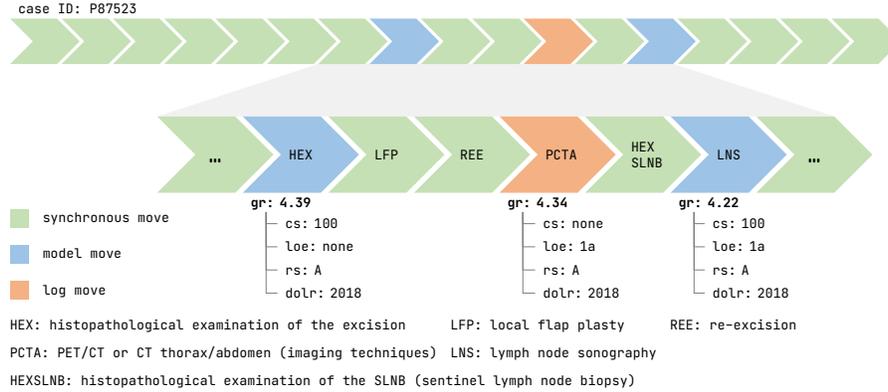
Since the guideline, according to its intention, mainly gives recommendations that have a higher degree of recommendation, a higher level of evidence, and a good consensus, the cost function was adjusted so that deviations from the optimum were weighted more heavily, this was ensured by squaring the weight term  $w_{c((s', s''))}$ . The computed fitness values for the five patients with the original approach [8] and the adapted weighted approach are shown in Table 1.

	violation weighted	non-weighted
Log Fitness	0.8642	0.8306
Fitness P87523	0.8787	0.8636
Fitness P56156	0.9258	0.8281
Fitness P21333	0.7840	0.8125
Fitness P23144	0.7769	0.8947
Fitness P23342	0.4444	0.3337

**Table 1.** Resulting fitness values compared with the non-weighting approach (log fitness and treatment trace for patients P21333-P87523).

The results show that the fitness values of the entire logs differ only slightly. This is due to the fact that most of the guideline recommendations have a high degree of recommendation. Moreover, not only the optimal alignments are weighted, but also the reference alignments. Thus, the fitness is averaged here as well. This is clearly visible in patient P23342. Here, two guideline recommendations were violated and three were evoked. Each of the recommendations has the highest level of evidence, the highest consensus strength, recommendation strength and the year 2018, as the year of the last review and thus is critical. This results in a weight of 2 for each recommendation, resulting in a fitness of 0.4444 for the weighted and 0.3337 for the unweighted approach. For patient

P87523 (see Figure 2), three MLMs are evoked and one (guideline recommendation 4.22) is critical. Since each weighting attribute has the highest weight, the deviation from recommendation 4.22 has a weight of 2.



**Fig. 2.** Aligned trace of the patient case P87523. Containing three moves in the aligned trace: two model moves and one log move. For each alignment step, the guideline recommendation (gr) is shown, which is incorporated in the respective MLM. Below that, the weighting attribute information for deriving the weights is shown.

The high fitness values close to the unweighted values show that the treatment traces in particular violate important statements. There are nearly no recommendation violations weighted as less important. In total, the traces violated 21 statements, of which 11 rules have a weight of 1 (as in a crisp approach) and 4 are critical with a weight of 2. In six violations, all for patient P56156 (11 violations in total), the weights are less than 1, with an average weight of 0.6.

## 5 Discussion

As demonstrated in Section 4, the weighted fitness measure provides little difference from the crisp approach when (1) there are few or no strongly weighted deviations (2) there are few deviations in general, and they are not sufficient to make a difference (3) in our approach, few MLMs are activated for treatment. Addressing this issue would require further investigation of the effect of the weights. An extended weighting scale could generate larger differences between individual results and better differentiate deviations in terms of their importance.

Furthermore, it must be considered that the creation and assignment of the weights and their levels is done manually. This implies a certain amount of effort, which usually requires the contribution of one or more domain experts for the corresponding case. In our evaluation, we were able to derive the weights from the medical classifications. However, if weights are to be implemented when there

is no default of importance, then they must be created based on the available data as well as a consensus of the respective domain experts. In addition, the assignment of numerical values to level of evidence and recommendation strength must be regarded critically, because it cannot be said with certainty that, e.g., the distance between *loe* 1a and 1b is the same as between 4 and 5.

The presented approach extends alignment-based conformance checking with weights to differentiate the severity of deviations. However, the data perspective is not currently considered, as it brings its own challenges, such as the semantically correct severity of a deviation from a given stage value.

When considering the results and the data set used, it should be noted that in an extended evaluation, the weighted fitness values may show greater differences from the unweighted fitness values. Since only a guideline section was modeled, only a delimited area of the entire treatment is tested for compliance. Accordingly, if a full treatment were reviewed, it is also very likely that more guideline violations of varying relevance would be identified, and the result would deviate much more significantly from the unweighted fitness score. Moreover, this work has shown that it is not straightforward to incorporate the importance of activities in the fitness value. On the one hand, the generated results could not show large differences in some cases and on the other hand, it is questionable to what extent fitness is the appropriate place to integrate the importance aspect. For medical process mining in particular, consideration should be given to introducing a new metric specifically designed for this purpose. In general, empirical research is needed on the association of greater guideline deviation and worse clinical outcomes addressed by clinical trials.

## 6 Conclusion

The presented approach for weighting violations of specific conditions allows the inclusion of attributes such as importance or soundness of modeled behavior. In the presented use case, this enables a more accurate knowledge representation in the process models and a higher expressiveness of the fitness value.

A limitation of the current approach is that it only considers the importance. However, the results show that the degree of deviation from the model is also important for calculating meaningful fitness values. This also applies to most cases of larger deviations in the time perspective since they should be weighted more heavily than small deviations. Accordingly, the replacement of one activity with another similar activity would also be less severe. An approach to include the degree of deviation for the data perspective could be the adaptation of the fuzzy set approach according to Zhang et al. [16]. Another factor could be the degree to which the conditions are met. Thus, it is interesting to know how close the trace could be to a threshold so that the respective condition still takes effect. Another challenge is the mapping of optional rules in the fitness value, which turned out to be very domain dependent. In future work, we intend to extend the approach to include the degree of deviation. In addition, the approach will be implemented and evaluated for several process modeling languages.

## References

1. Leitlinienprogramm Onkologie (Deutsche Krebsgesellschaft, Deutsche Krebshilfe, AWMF): Diagnostik, Therapie und Nachsorge des Melanoms, Langversion 3.3, 2020, AWMF Registernummer: 032/024OL, <http://leitlinienprogramm-onkologie.de/leitlinien/melanom/>, (accessed 20.09.2022)
2. van der Aalst, W., Adriansyah, A., van Dongen, B.: Replaying history on process models for conformance checking and performance analysis. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **2**(2), 182–192 (2012)
3. van der Aalst, W.M.P.: *Process Mining: Data Science in Action*. Springer Berlin Heidelberg, 2nd edn. (2016)
4. Arkad, K., Gill, H., Ludwigs, U., Shahsavar, N., Gao, X.M., Wigertz, O.: Medical logic module (mlm) representation of knowledge in a ventilator treatment advisory system. *International journal of clinical monitoring and computing* **8**(1) (1991)
5. Dunzer, S., Stierle, M., Matzner, M., Baier, S.: Conformance checking: a state-of-the-art literature review. In: Betz, S. (ed.) *Proceedings of the 11th International Conference on Subject-Oriented Business Process Management*. pp. 1–10. Association for Computing Machinery, New York (2019)
6. Graham, R., Mancher, M., Miller Wolman, D., Greenfield, S., Steinberg, E. (eds.): *Clinical practice guidelines we can trust*. National Academies Press (2011)
7. Grüger, J., Geyer, T., Bergmann, R., Braun, S.A.: CGK4PM: Towards a methodology for the systematic generation of clinical guideline process models and the utilization of conformance checking. *BioMedInformatics* **2**(3), 359–374 (2022)
8. Grüger, J., Geyer, T., Kuhn, M., Bergmann, R., Braun, S.A.: Declarative guideline conformance checking of clinical treatments. In: *Business Process Management Workshops*. Preprint at <https://arxiv.org/abs/2209.09535> (2022)
9. Grüger, J., Geyer, T., Kuhn, M., Braun, S.A., Bergmann, R.: Verifying guideline compliance in clinical treatment using multi-perspective conformance checking: A case study. In: *Process Mining Workshops*, vol. 433, pp. 301–313. Springer International Publishing and Imprint Springer, Cham (2022)
10. Hripcsak, G.: Writing Arden Syntax medical logic modules. *Computers in Biology and Medicine* **24**(5), 331–363 (1994)
11. Lohr, K.N., Field, M.J.: *Clinical practice guidelines: Directions for a new program*, Publication IOM, vol. 90-08. National Academy Press, Washington (1990)
12. Munoz-Gama, J., et al.: Process mining for healthcare: Characteristics and challenges. *Journal of Biomedical Informatics* **127**, 103994 (2022)
13. Oliart, E., Rojas, E., Capurro, D.: Are we ready for conformance checking in healthcare? measuring adherence to clinical guidelines: A scoping systematic literature review. *Journal of Biomedical Informatics* **130**, 104076 (2022)
14. Samwald, M., Fehre, K., de Bruin, J., Adlassnig, K.P.: The arden syntax standard for clinical decision support: experiences and directions. *Journal of biomedical informatics* **45**(4), 711–718 (2012)
15. van Zelst, S.J., Mannhardt, F., de Leoni, M., Koschmider, A.: Event abstraction in process mining: literature review and taxonomy. *Granular Computing* **6**(3), 719–736 (May 2020)
16. Zhang, S., Genga, L., Yan, H., Lu, X., Duan, H., Kaymak, U.: Towards multi-perspective conformance checking with fuzzy sets (2020), <https://arxiv.org/abs/2001.10730>