

RESEARCH

Open Access



# Acquisition of temporal patterns from electronic health records: an application to multimorbid patients

Alicia Ageno<sup>1,2\*</sup>, Neus Català<sup>1,2</sup> and Marcel Pons<sup>3</sup>

## Abstract

**Background** The exponential growth of digital healthcare data is fueling the development of Knowledge Discovery in Databases (KDD). Extracting temporal relationships between medical events is essential to reveal hidden patterns that can help physicians find optimal treatments, diagnose illnesses, detect drug adverse reactions, and more. This paper presents an approach for the extraction of patient evolution patterns from electronic health records written in Catalan and/or Spanish.

**Methods** We propose a robust formulation for extracting Temporal Association Rules (TARs) that goes beyond simple rule extraction by considering the sequence of multiple visits. Our highly configurable algorithm leverages this formulation to extract Temporal Association Rules from sequences of medical instances. We can generate rules in the desired format, content, and temporal factors while accounting for different levels of abstraction of medical instances. To demonstrate the effectiveness of our methodology, we applied it to extract patient evolution patterns from clinical histories of multimorbid patients suffering from heart disease and stroke who visited Primary Care Centers (CAP) in Catalonia. Our main objective is to uncover complex rules with multiple temporal steps, that comprise a set of medical instances.

**Results** As we are working with real-world, error-prone data, we propose a process of validation of the results by expert practitioners in primary care. Despite our limited dataset, the high percentage of patterns deemed correct and relevant by the experts is promising. The insights gained from these patterns can inform preventive measures and help detect risk factors, ultimately leading to better treatments and outcomes for patients.

**Conclusion** Our algorithm successfully extracted a set of meaningful and relevant temporal patterns, especially for the specific type of multimorbid patients considered. These patterns were evaluated by experts and demonstrated the ability to predict risk factors that are commonly associated with certain diseases. Moreover, the average time gap between the occurrence of medical events provided critical insight into the term of these risk factors. This information holds significant value in the context of primary healthcare and preventive medicine, highlighting the potential of our method to serve as a valuable medical tool.

**Keywords** Electronic health records, Temporal data mining, Temporal association rules, Clinical decision support systems, Risk factors detection

\*Correspondence:

Alicia Ageno  
ageno@cs.upc.edu

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

The increasing adoption of Electronic Health Records (EHRs) in healthcare systems offers the potential for leveraging large amounts of health data to improve the efficiency and accuracy of medical practitioners. To support clinical diagnosis, prevention, and administrative decision-making, healthcare professionals can benefit from data mining technologies [1] and/or machine learning techniques [2]. These techniques can be used to model information from multiple data sources, such as images [3, 4], DNA sequences [5] or structured and unstructured textual data. In particular, automatic analysis of EHRs can be essential for identifying patterns, preventing errors, improving quality, reducing costs, and saving time for healthcare services. Several studies have already demonstrated the potential of data mining techniques to extract valuable insights from EHRs [6, 7].

In the context of medical diagnosis, understanding the interaction between medical instances during a patient visit is crucial for early detection and prevention of new diseases. However, research on the temporal relationships between these instances is lacking. Therefore, our goal in this work is to explore the possibility of modelling patient evolution patterns based on their clinical histories, to identify interesting associations such as drug side effects, recurrent chronic symptoms, and co-occurring diagnoses. By doing so, we hope to improve the medical tools available to physicians and ultimately provide more personalized patient care.

EHRs are essential sources of temporal information that provide a comprehensive understanding of healthcare events. To extract implicit, non-trivial, and potentially useful abstract information from large collections of temporal data, temporal data mining techniques have been explored [8]. Our approach focuses on one of these techniques, namely temporal association rules [9–16]. We aim to develop a new variant of temporal association rule that incorporates our formulae for computing the associated measures of support and confidence. Furthermore, we will propose two variants of a new algorithm for mining these temporal association rules.

Although our framework will be general, in this article we will focus on its application to knowledge discovery for risk prediction and disease diagnosis in multimorbid patients. These patients are often affected by multiple disorders and symptomatology, making them a prevalent issue in certain clinical contexts, such as primary care. Despite their prevalence, the treatment of multimorbid patients has not been extensively studied. To address this research gap, we aim to develop a robust framework for extracting temporal association rules from multi-attribute data.

Temporal association rules are a type of sequential pattern mining that aims to identify associations between events that occur in a particular temporal order. By analyzing the timing and sequence of events, temporal association rules can provide valuable insights into the underlying relationships between different health conditions or symptoms. Our approach will focus on the temporal aspect of the data, which will be used as an input parameter and obtained as part of the results. We will also incorporate the ability to express healthcare instances at different levels of abstraction, allowing us to extract patterns with different degrees of generalization<sup>1</sup>. Our ultimate goal is to develop a comprehensive framework for extracting these rules from multi-attribute data in order to improve our understanding of the complex relationships between health conditions.

Association rule mining is a data mining technique that aims to discover relationships, patterns, or rules among variables in a dataset. The origins of association rule mining can be traced back to the Apriori algorithm [17]. Since then, several extensions have been developed for different purposes [18], such as sequential pattern mining [19] which aims to identify patterns of ordered events in a dataset by finding frequent sequences of itemsets. An example of such an extension is the Generalized Sequential Pattern (GSP) algorithm [20]. The problem of learning rules from time series data, where the antecedent tends to appear in conjunction with the consequent according to certain time constraints, is called *temporal association rule mining*.

A temporal association rule (TAR) is a type of temporal relationship that describes a connection between an antecedent and a consequent. Although temporal pattern mining in general (and temporal association rule mining in particular) has received considerable attention in recent years, one of the main problems in this area is the lack of visibility of most of the work, as there is no standard terminology to refer to. Different formulations of TARs can be found in the literature, each with different approaches to computing support and confidence, and each introducing the temporal component in different ways, making it difficult to find and compare proposals and studies in this area.

Mooney and Roddick [21] describes several sequential pattern mining approaches, including temporal sequences and more specifically TARs. These approaches extract patterns with only one antecedent and one consequent. Eventually, the antecedent may consist of several sequential elements, but there is no

---

<sup>1</sup> Throughout this work, we will use the terms patterns and rules interchangeably to refer to the concept of temporal association rules.

way to detail temporal constraints among them. However, our algorithm has been designed to extract rules with different numbers of temporal steps (allowing a different interval [t1, t2] for each step), attributes and levels of concept abstraction.

Furthermore, a comprehensive overview of the different methods for temporal association rule mining is provided in [9]. The work proposes a taxonomy in which they classify the existing approaches at a first level into two categories, based on the use of the time variable: (1) “time as implied component”, those that use temporally ordered datasets to discover temporal constraints, and (2) “time as integral component” those that integrate the time variable as another attribute of the data and analyze the temporal aspects in which the rules occur. According to their definition, our methodology would be a hybrid belonging to both first-level categories, “time as implied” (with second-level category “sequential”) and “time as integral component” (with second-level category “time-interval”). This is so because, on the one hand, the time variable provides order and temporal constraints to the EHR database, and on the other hand it is also used as a parameter in the process of pattern extraction (in the form of time gaps). The article also describes a large number of applications in the areas of medicine and healthcare (such as [22–25]). However, none of them is directly comparable to our approach, as they only belong to one of the categories in the taxonomy.

In [10] and [11], an algorithm for the extraction of temporal association rules is presented, which deals with complex temporal patterns, represented through the formalism of Temporal Abstractions. This algorithm allows dealing both with events with a certain time duration and with events that occur at a single moment in time.

In their paper, Zhan et al. [12] present a more generalized form of temporal association rules. These extended temporal association rules take the form of an implication  $X_1 \xrightarrow{t_1} X_2 \xrightarrow{t_2} \dots \xrightarrow{t_{p-1}} X_p$  with  $p \geq 2$ , where each  $t_i$  is a time constraint. In addition, a modified formulation for computing support and confidence for these rules is defined. Furthermore, the authors introduce another form of temporal association rules,  $X_1 \xrightarrow{T_1} X_2 \xrightarrow{T_2} \dots \xrightarrow{T_{p-1}} X_p$ , where  $T_i = [t_{i1}, t_{i2}]$  are time intervals with  $t_{i1} < t_{i2}$ . They use these definitions to propose a fast algorithm for mining temporal association rules, which they test on both synthetic and real datasets, demonstrating excellent performance. Our work is based on a combination of ideas from [10] and [12]. By leveraging these insights, we aim to advance the application of temporal association rules and develop more effective methods for knowledge discovery.

## Methods

### Electronic health record (EHR) dataset

Our dataset contains information on patient visits at Primary Care Centers (CAP) in Catalonia. This dataset has been provided by IDIAP JGol [26] and consists of EHRs of multimorbid patients visited between 2010 and 2016. In our case, a multimorbid patient is a patient whose medical history includes at least one of the following symptoms, which must have occurred when the patient was over 50 years of age:

- Transient ischaemic attack
- Haemorrhagic strokes
- Ischaemic strokes
- Other types of strokes
- Stroke-related sequelae
- Lung cancer
- Colorectal cancer
- Acute myocardial infarction
- Exitus

Our EHRs, written in Catalan and/or Spanish, are composed of four sections: reason for the consultation, medical examination, evaluation and medical treatment plan. These EHRs were manually annotated and represented in a graph format (Fig. 1) with four types of nodes: *Body part*, *Diagnosis*, *Drug*, and *Sign or symptom*. These nodes will represent our healthcare instances. In addition, six types of relationship were annotated: *before*, *causality\_of*, *coOccur*, *cotreated\_with*, *located\_in*, and *substituted\_by*. The “Appendix” contains a detailed description of node properties (or attributes) and relationships. As the corpus has been manually annotated, it may contain some inconsistencies. Indeed, the manual generation of high quality annotated medical records for such complex concepts and relationships has been particularly challenging, highlighting the need for further development of resources for annotating EHRs.

The IDIAP dataset contains information about 320 patients making a total of 72,257 healthcare instances, distributed as shown in Table 1. We focus on two node attributes, the *code* associated with each node and the *raw\_text*, which is the description that the practitioner entered for each medical instance. The table shows the number of different codes, the number of different text descriptions, and the total number of instances for each type of node. There are a total of 14,330 visits. Each patient attended an average of 44.78 visits during these six years, with a standard deviation of 34.83. The patient with the most visits had 205 visits and the patient with the fewest visits had only 1 visit. Visits have an average of 5.04 healthcare instances. The mean time between two consecutive visits for a patient is 42.83 days and the

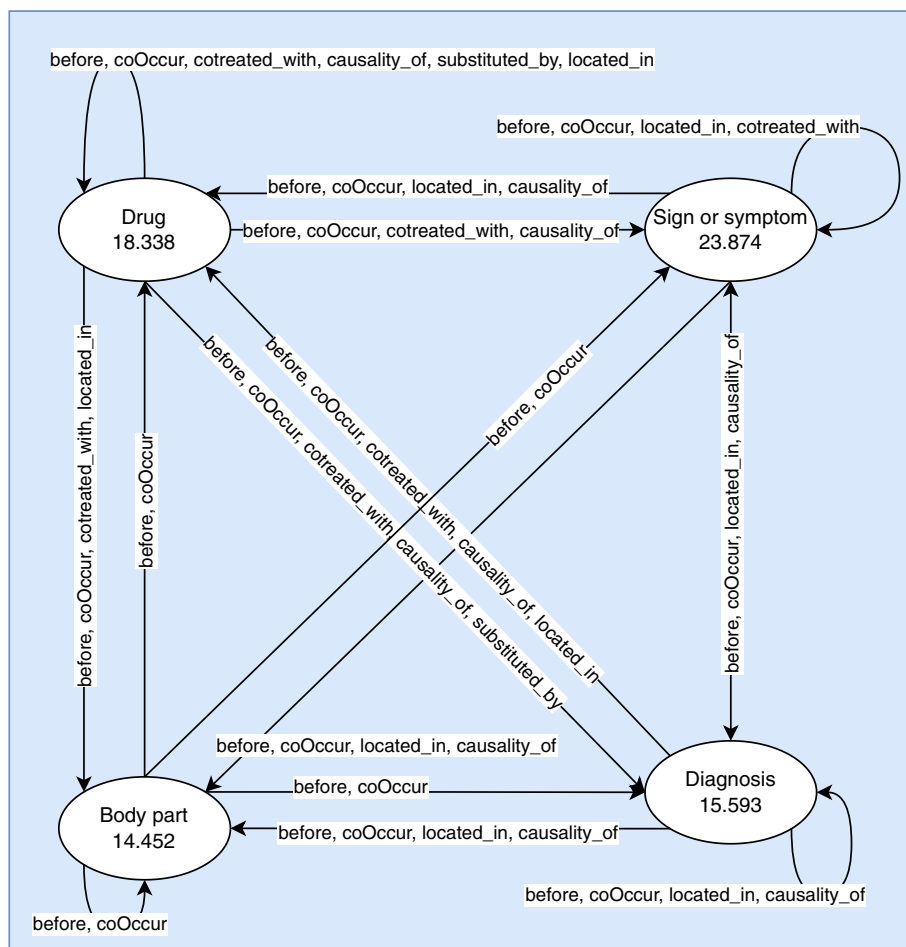


Fig. 1 Data model of IDIAP dataset

**Table 1** Distribution of healthcare instances and their attributes before preprocessing

Node type	Codes	Raw_texts	Instances
Body part	601	5184	14452
Diagnosis	1193	6429	15593
Drug	796	3742	18338
Sign or symptom	249	10455	23874
Total	2839	25810	72257

standard deviation is 86.29 days, with a minimum of 1 day and a maximum of 2,005 days (5.5 years). These annotated EHRs have been imported into a Neo4J [27] database, from which the information can be extracted.

**Data preparation**

The first stage of data preprocessing consisted of the filtering of the annotations according to the *time* and

*certainty* attributes (see “Appendix”). The following preprocessing steps were applied to the *raw\_text* attribute: case folding, stopword removal and string distance. The latter means that for those *raw\_text* values that have more than three characters, we grouped all the values that had a Levenshtein distance equal to or less than two. Furthermore, some mistyped codes were corrected.

Table 2 shows the number of different values for the attributes *code* and *raw\_text* after preprocessing. Despite this preprocessing, we can see that our dataset is quite limited. The number of different healthcare instances is still large with respect to the number of visits, so the frequency of occurrence of each medical instance is small. This may imply low supports for the associations, which we will try to increase through generalization.

Different levels of generalization can be applied to each type of *code*, except for the *Body part* nodes (see “Appendix” for a description of the levels). The use of *raw\_text* was discarded because the small size of the dataset did not allow sufficient generalization. Table 2 shows the high

**Table 2** Distribution of healthcare instances and their attributes after preprocessing

Node type	Codes	Raw_texts	Instances
Body part	601	4103	14452
Diagnosis	1158	4003	15012
Drug	792	1930	18086
Sign or symptom	225	8138	22835
Total	2776	18174	70385

variability of the raw texts compared to the number of occurrences. In this work, we have decided to use a level of generalization where we use level 2 for *Diagnosis*, level 5 for *Drug* and level 2 for *Sign or symptom*. We consider that these levels provide a good compromise between interpretability and generalization. When we refer to a code level, we specify the maximum code level allowed. However, if a given node does not provide the specified level, its maximum level of generalization will be used.

**Temporal association rules formulation**

In this section, we will present the formulation for computing our temporal association rules. As mentioned above, this formulation is an extension of the one defined in [12].

Let  $\mathcal{J}$  be a set of healthcare instances, denoted by  $I_1, I_2, \dots, I_m$ . Let  $\mathcal{D}$  be a set of temporal transactions, where each transaction  $(S_i, t_i)$  corresponds to a healthcare visit  $S_i$  attended at time  $t_i$ . First, we define temporal association rules of the form  $X \xrightarrow{t} Y$ , where  $X \subset \mathcal{J}$ ,  $Y \subset \mathcal{J}$  and  $t$  is a unit of time (in our context, days). These rules imply that a time gap of  $t$  has elapsed between an occurrence of instances  $X$  and an occurrence of instances  $Y$ . Unlike conventional association rules,  $X \cap Y$  need not necessarily be  $\emptyset$ , since it makes sense for the same healthcare events to continue to occur after some time.

Let

$$g_{S_i}^p : \mathcal{P}(\mathcal{J}) \rightarrow \{0, 1\}$$

$$g_{S_i}^p(X) = \begin{cases} 1 & \text{if } S_i \text{ exists and contains } X \text{ for patient } p \\ 0 & \text{else} \end{cases}$$

where  $S_i$  is a visit at time (day)  $i$ . Then

$$g_{S_i,t}^p : \mathcal{P}(\mathcal{J})^2 \rightarrow \{0, 1\}$$

$$g_{S_i,t}^p(X, Y) = \min(g_{S_i}^p(X), g_{S_{i+t}}^p(Y))$$

where  $p$  is a patient,  $S_i$  is a visit at time  $i$ , and  $t$  is a time interval (or temporal gap) measured in days. This function returns a value of 1 if, at time  $i$ , the patient  $p$  has healthcare instances  $X$ , and at time  $i + t$ , the same patient  $p$  has instances  $Y$ . Otherwise, the function returns a value

of 0. In our formulation, the parameter  $\emptyset$  simply indicates the occurrence of a visit at the given time. For example,  $g_{S_i,t}^p(X, \emptyset)$  is equal to 1 if, for a given patient  $p$ ,  $S_i$  exists and contains  $X$ , and there is a visit  $S_{i+t}$ .

Let

$$h_t : \mathcal{P}(\mathcal{J})^2 \rightarrow \mathbb{N}, \quad h_t(X, Y) = \sum_p \sum_{S_i \in \mathcal{D}} g_{S_i,t}^p(X, Y)$$

$h_t$  is the count support for  $X, Y$ . The inner sum accounts for the number of times the patient  $p$  presents  $X$  at time  $i$  and  $Y$  at time  $i + t$ . The outer sum is done to take into account all the patients. This formula is used to compute the support and confidence of temporal association rules  $X \xrightarrow{t} Y$ .

Furthermore, we extend these temporal association rules to deal with time intervals. These rules are expressed as  $X \xrightarrow{T} Y$ , where  $T = [t_1, t_2] (t_1 < t_2)$ . We denote  $t_1$  as *min\_gap* and  $t_2$  as *max\_gap*. Then:

$$h_T : \mathcal{P}(\mathcal{J})^2 \rightarrow \mathbb{N}$$

$$h_T(X, Y) = \sum_p \sum_{S_i \in \mathcal{D}} \min \left( \sum_{t \in T} g_{S_i,t}^p(X, Y), 1 \right)$$

$h_T$  is the count support for  $X, Y$ . The inner sum accounts for the number of times that a patient  $p$  presents  $X$  at time  $i$  and  $Y$  at time  $i + t$  in a set of visits  $S_i$ . The *min* function ensures that, in case the association  $X \xrightarrow{T} Y$  exists, it is counted only once for each consequent. The summation  $\sum_{S_i \in \mathcal{D}}$  takes into account all visits of a patient, while the outer summation accounts for all patients. Using these formulae, we can define our version of the following measures:

$$support(X \xrightarrow{T} Y) = \frac{h_T(X, Y)}{h_T(\emptyset, \emptyset)} \tag{1}$$

$$confidence(X \xrightarrow{T} Y) = \frac{h_T(X, Y)}{h_T(X, \emptyset)} \tag{2}$$

$$rulesRespectConsequent(X \xrightarrow{T} Y) = \frac{h_T(X, Y)}{h_T(\emptyset, Y)} \tag{3}$$

$$lift(X \xrightarrow{T} Y) = \frac{support(X, Y)}{support(X, \emptyset) \cdot support(\emptyset, Y)} \tag{4}$$

$$leverage(X \xrightarrow{T} Y) = support(X, Y) - support(X, \emptyset) \cdot support(\emptyset, Y) \tag{5}$$

In particular, *rulesRespectConsequent* (Eq. 3) is a new measure requested by the experts during the evaluation

phase (see “Results” section). It denotes the number of medical cases that feature the antecedents in the elapsed gap, out of all the cases that present the consequent. The rest are widely known measures.

When applying the previous formulation to automatically infer patterns that help to understand the clinical conditions of multimorbid patients, we faced a dilemma. On the one hand, temporal association rules could reflect recurrent patterns presented by the patients, extracted as the repetition of medical events frequently presented by any patient. The rationale of this implementation is consistent with the conventional Apriori algorithm, which does not distinguish the agent presenting the patterns, so this approach will be referred to as *apriori-like*. On the other hand, temporal association rules may reflect patterns that frequently occur among patients, ignoring the repetition of healthcare events within each patient. This approach aims to generate common rules that are present in the clinical data and might be more useful for providing information about conventional (non-multimorbid) patients, so it will be referred to as *patient-oriented*. The previous formulation needs to be adapted to these two approaches, though it is not included here to make the text more readable.

The formulations presented here can be extended to cases where a sequence of antecedents or temporal steps leads to a given consequent. The resulting temporal association rules have the form  $X_1 \xrightarrow{T_1} X_2 \xrightarrow{T_2} \dots \xrightarrow{T_n} X_n$ , where each  $X$  represents either a single instance of a diagnosis, drug, body part, symptom or sign or a combination of these. In this notation, the arrows represent a time interval between each sequence of events, or different healthcare visits occurring at different times. These temporal steps can be time specific ( $t$ ) or intervals ( $T$ ).

The resulting formulation is a modification of the one presented in [12]. Let  $t = (t_1, \dots, t_{q-1})$

$$g_{S_i,t}^p : \mathcal{P}(\mathcal{J})^q \rightarrow \{0, 1\}$$

$$g_{S_i,t}^p(X_1, X_2, \dots, X_q) = \min\left(g_{S_i}^p(X_1), g_{S_{i+t_1}}^p(X_2), \dots, g_{S_{i+t_1+\dots+t_{q-1}}}^p(X_q)\right)$$

The computation of the support count function changes according to the version. For the *apriori-like* approach:

$$h_t^p : \mathcal{P}(\mathcal{J})^q \rightarrow \mathbb{N}$$

$$h_t^p(X_1, X_2, \dots, X_q) = \sum_{S_i \in D} g_{S_i,t}^p(X_1, X_2, \dots, X_q) \tag{6}$$

And for the *patient-oriented* approach:

$$h_t^p : \mathcal{P}(\mathcal{J})^q \rightarrow \mathbb{N}$$

$$h_t^p(X_1, X_2, \dots, X_q) = \min\left(\sum_{S_i \in D} g_{S_i,t}^p(X_1, X_2, \dots, X_q), 1\right) \tag{7}$$

Summing up for all the patients:

$$h_t(X_1, X_2, \dots, X_q) = \sum_P h_t^p(X_1, X_2, \dots, X_q) \tag{8}$$

Then, the rule indicators can be computed with the same formulae for both versions:

$$\text{support}(X_1 \xrightarrow{t_1} X_2 \dots \xrightarrow{t_{q-1}} X_q) = \frac{h_t(X_1, X_2, \dots, X_q)}{h_t(\emptyset, \dots, \emptyset)}$$

$$\text{confidence}(X_1 \xrightarrow{t_1} X_2 \dots \xrightarrow{t_{q-1}} X_q) = \frac{h_t(X_1, X_2, \dots, X_q)}{h_t(X_1, X_2, \dots, X_{q-1}, \emptyset)}$$

$$\text{rulesRespectConsequent}(X_1 \xrightarrow{t_1} X_2 \dots \xrightarrow{t_{q-1}} X_q) = \frac{h_t(X_1, X_2, \dots, X_q)}{h_t(\emptyset, \dots, X_q)}$$

$$\text{lift}(X_1 \xrightarrow{t_1} X_2 \dots \xrightarrow{t_{q-1}} X_q) = \frac{h_t(X_1, X_2, \dots, X_q)}{h_t(X_1, X_2, \dots, X_{q-1}, \emptyset) * h_t(\emptyset, \dots, X_q)}$$

$$\text{leverage}(X_1 \xrightarrow{t_1} X_2 \dots \xrightarrow{t_{q-1}} X_q) = h_t(X_1, X_2, \dots, X_q) - h_t(X_1, X_2, \dots, X_{q-1}, \emptyset) * h_t(\emptyset, \dots, X_q)$$

The previous formulation is also extended to deal with time intervals, although we will not go into the details here.

### Temporal association rules extraction

Using the previous formulation, we have developed two algorithms for mining temporal association rules using both *apriori-like* and *patient-oriented* approaches. These algorithms are based on the work of [10] and [12]. Both algorithms leverage pre-computed data structures based on the generalization levels of healthcare instances. These algorithms also take as input the minimum support and confidence required (*minsup* and *minconf*) and a list of  $n$  time intervals  $[T_1, T_2 \dots T_n]$  whose length represents the maximum number of temporal steps to be considered for the rules.

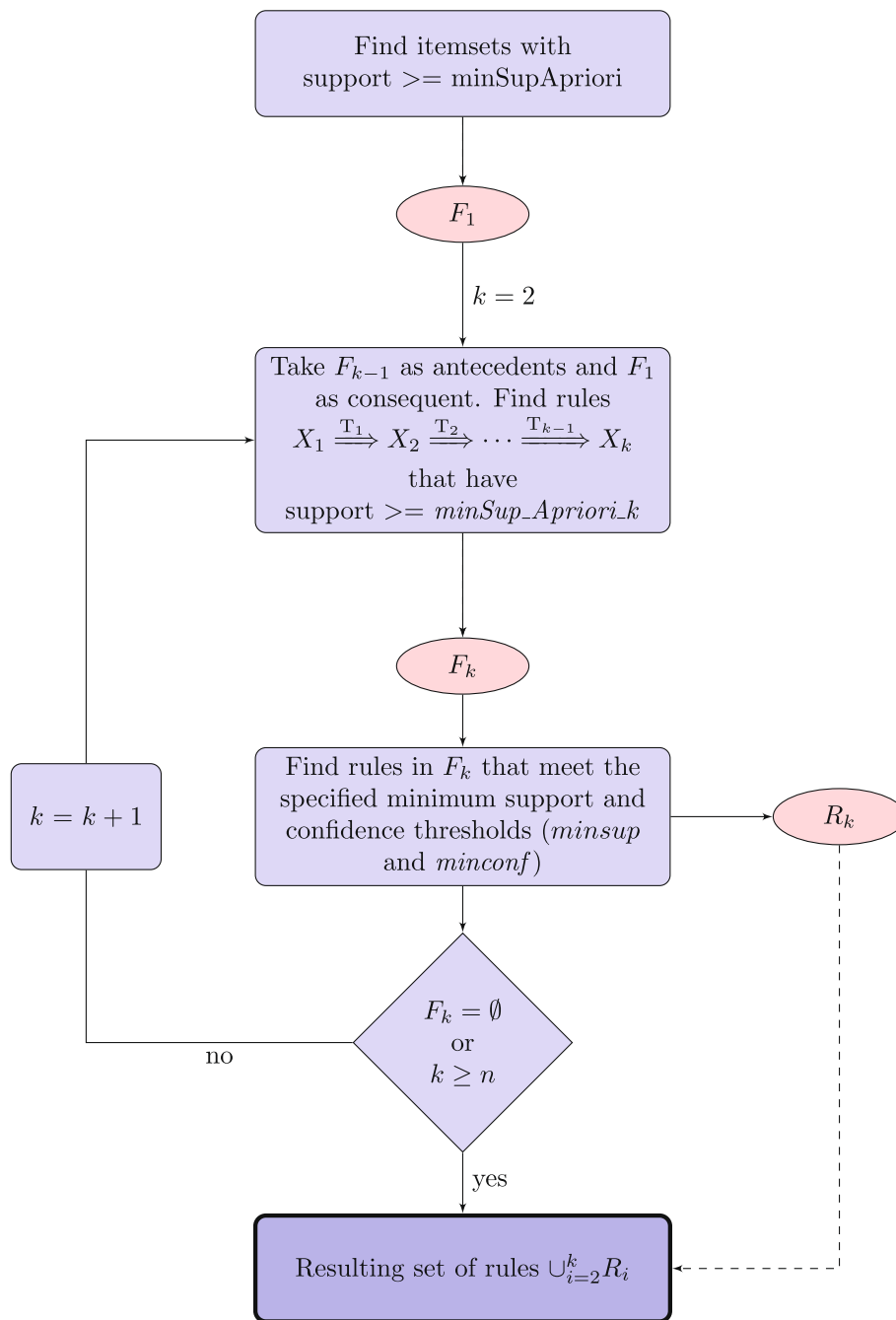
Both algorithms share a common structure, but the main difference between them lies in the counting methodology. The *apriori-like* approach counts all consecutive events that satisfy temporal constraints, while the *patient-oriented* approach counts patients that satisfy temporal constraints across their visits. In particular, if the same set of consecutive events satisfies the condition  $a_i \xrightarrow{T_1} a_j$  in one patient multiple times, it is counted only once. The differences between the two algorithms are implemented in the code by calculating support and confidence differently, as described in the formulations. In addition, a new variable, *minSup\_Apriori*, is

introduced to take into account the minimum support of the previous time steps. This variable is a percentage value. Figure 2 provides a brief description of the steps involved in the algorithm.

The algorithms return a list of rules along with their corresponding metrics: *support*, *confidence*, *lift*, *leverage*, and *rulesRespectConsequent*. Furthermore, the average time gap between each successive event of the rules is computed. For *apriori-like*, the average

time gap is calculated by taking the average of all the time gaps that meet the thresholds for the specific step. For *patient-oriented*, it is calculated by taking the average of a list, where each item is the average of the time gaps satisfying the thresholds presented by a patient.

By providing all these parameters, we aimed to design a highly adaptable code, not only in terms of *support* and *confidence* measures but also to allow the extraction of



**Fig. 2** Flow diagram of the temporal association rule mining algorithm. Pink ovals represent intermediate item sets ( $F_1$ ) and sets of rules ( $F_k$  and  $R_k$ )

rules with different numbers of temporal steps, different time gaps and different levels of abstraction for each instance. The aim is to provide a powerful tool for generating rule sets that can be applied across various domains and applications. The source code of the algorithms is available at [28].

## Results and discussion

A good diagnosis must take into account the patient's medical history. As mentioned in the "Background" section, we have tested our framework mainly to obtain patterns to help in diagnosis or to detect risk factors that may somehow derive certain diseases. Since there is no gold standard of valid patterns, the evaluation of the obtained results in clinical practice is usually based on the interpretation by domain experts, who may consider various criteria to assess their quality, relevance, and clinical significance. It's important to note that these criteria may vary depending on the specific clinical context, disease area, or research objectives.

In our work, the evaluation of the generated patterns was carried out by three experts, experienced primary care physicians from IDIAP [26]. As the usefulness of a pattern ultimately depends on the experts' ability to apply it or integrate it into the decision-making process, we conducted initial test evaluations using a small set of patterns. After the preliminary evaluation process, the authors and the experts agreed on two specific criteria:

- **Correctness:** Experts assess whether the patterns align with known clinical knowledge, established medical guidelines, or domain expertise. Patterns are intuitive and logically consistent. Experts consider whether the patterns are easy to understand and provide clear insights into patient characteristics, disease trajectories, risk factors, treatment effectiveness or other relevant clinical factors. They also assess whether the patterns can effectively predict future events, disease progression, treatment response, or adverse outcomes.
- **Relevance:** Experts assess whether the patterns reflect meaningful associations or dependencies that have practical utility for patient care, disease progression, treatment response or other clinical outcomes. They assess whether the patterns can contribute to improved diagnosis, prognosis, treatment selection, risk assessment or patient monitoring. Therefore, experts may consider a pattern that does not represent known medical knowledge to be relevant.

We then worked together to develop a set of labels to assess both criteria. These labels allow us to determine the level of agreement among experts on our findings. The labels for *Correctness* evaluation can be defined as follows:

- **Totally incorrect:** The rule is completely wrong or contradicts medical knowledge.
- **Totally correct:** The rule is completely right and aligns with medical knowledge.
- **Partially correct (temporal facet):** The rule is correct, but any of the temporal gaps between antecedents and consequent are not the expected ones.
- **Partially correct (clinical facet):** The rule is partially correct due to the absence or excess of any antecedent.
- **Partially correct (both facets):** The rule is partially correct due to the absence or excess of any antecedent and the inappropriateness of any of the time gaps.

In terms of *Relevance*, the rules have been sorted according to the following classes:

- **Not relevant:** The rule is considered worthless (it presents no significant clinical relevance).
- **Relevant and known:** The rule has significant clinical relevance, and it is a well-known implication (common knowledge among experts).
- **Relevant and unknown:** The rule has significant clinical relevance and should be studied further, as it is not a well-known implication.

It is important to emphasize that we are not looking for strict cause-effect relationships here but for patterns of behaviour. This means that there can be any kind of relationship between antecedents and consequent.

We generated several packs of rules, each representing different parameterizations. Due to the large number of combinations of parameter values, we fixed some of them and focused on testing both approaches (*apriori-like* and *patient-oriented*) under the following time constraints, which were agreed upon by the experts. For simplicity, the same time constraints were applied to each of the several possible temporal steps of the rules:

- **Shortest-term rules:** medical events that take place within 1 to 10 days (39% of the visits in our test set occurred within this interval).
- **Short-term rules:** medical events that occur within 5 to 30 days (45% of the visits).
- **Mid-term rules:** medical events that take place within 20 to 90 days (33% of the visits).



- Long-term rules: medical events that occur within 60 to 120 days (11% of the visits).

These packs of rules were individually annotated by the experts using the previous labels. Then they came to a common agreement. The experts have emphasized that the relevance of the patterns depends heavily on the time gaps. It is very different to be diagnosed with a cold than it is to be diagnosed with a malignant neoplasm. If a patient has a cold, it may be relevant to know what happened 3-15 days before, but in the case of malignant neoplasm, we would like to consider what happened, for instance, in the last 30-300 days.

Nevertheless, these executions still produce hundreds of rules, depending on the values of the parameters *min\_sup* and *min\_conf*. To filter these rules for each *min\_gap* - *max\_gap* parameterization, the first selection criterion was that they had to contain more than one temporal step, since having multiple temporal steps is essential for testing our methodology. These resulting multi-step rules were sorted according to *lift* and then *rulesRespectConsequent* values, and the top 20 were selected<sup>2</sup>. *Lift* measures the ratio of the confidence of the rule to the expected confidence of the rule (Eq. (45)). Overall, *lift* quantifies the degree of association between the elements on the antecedents and the consequent of a rule, and higher values of *lift* indicate stronger association. In this work, the sorting criterion for the rules was chosen on the basis of the *lift*, and in the case of ties, the frequency of occurrence would prevail.

The minimum support for each pack was adjusted based on the chosen approach and time gap. Initially, we started with values around 5% and gradually decreased them to obtain rules with more temporal steps. For instance, the minimum *minSup* was set to 0.5% to generate 3-step rules for the shortest-term time gap of the *apriori-like* approach. As for *minConf*, it was set at a high level with a minimum value of 80%.

After each pack was independently annotated by the three experts, we assessed the agreement achieved. Since the classification categories are not ordinal and three annotators were involved, we decided to use Fleiss' Kappa measure [29]. The results obtained (see Table 3), highlight the subjectivity and difficulty of the evaluation process. According to the standard interpretation [30], positive values below 0.2 would be considered as a "slight agreement". However, we would have expected a higher

**Table 3** Fleiss' Kappa agreement among experts

Approach	Correctness	Relevance
<i>apriori-like</i>	0.105	0.038
<i>patient-oriented</i>	0.119	0.002

**Table 4** Summary of *Correctness* results after consensus

	Correctness				
	Incorrect	Correct	Par-Temp	Par-Clinic	Par-Both
<i>apriori-like</i>					
Shortest-term	25	65	10	0	0
Short-term	35	60	5	0	0
Mid-term	50	45	5	0	0
Long-term	10	80	10	0	0
Global	30	62.5	7.5	0	0
<i>patient-oriented</i>					
Shortest-term	50	45	0	5	0
Short-term	45	35	10	10	0
Mid-term	45	35	10	10	0
Long-term	43.3	23.3	16.6	10	6.6
Global	45.82	34.57	9.15	8.75	1.65

(at least "fair") agreement, that would have indicated that these patterns corresponded to common medical knowledge among physicians. There may be two reasons for these figures. On the one hand, the small volume of data might produce patterns that reproduce less frequent phenomena (uncommon behaviour in the experts' knowledge), which may justify the inconsistency among the experts. On the other hand, the experts complained that the codes were too general, which made the evaluation harder. They also pointed out that, in many cases, healthcare instances are findings that you discover when performing other examinations. This is coherent with our interpretation of our rules as patterns of behaviour rather than strict cause-effect relationships.

Tables 4 and 5 show respectively a summary of the percentages of *Correctness* and *Relevance* in the annotated packs after the expert consensus. The results depend strongly on the approach. For *apriori-like*, regarding *Correctness*, we can see that two-thirds of the patterns are considered correct, indicating that our algorithm is extracting sound data. However, the *patient-oriented* approach performs worse. One explanation for this positive tendency towards the *apriori-like* method stems from the dataset used for the pattern extraction, which consisted of multimorbid patients whose visits present recurrent medical conditions and are therefore more suitable for *apriori-like*

<sup>2</sup> In addition, those rules containing sign or symptom with ICD-10 code AS99 ('Other specified general symptoms, complaints and abnormal findings') and rules containing ICD-10 code Z76.8 ('Persons encountering health services in other specified circumstances') were filtered for being too general.

**Table 5** Summary of *Relevance* results after consensus

	Relevance		
	No-Rel	Rel-Known	Rel-Unknown
<i>apriori-like</i>			
Shortest-term	35	65	0
Short-term	70	30	0
Mid-term	70	30	0
Long-term	10	90	0
Global	46.25	53.75	0
<i>patient-oriented</i>			
Shortest-term	50	50	0
Short-term	60	40	0
Mid-term	45	55	0
Long-term	63.3	36.6	0
Global	64.57	35.4	0

characteristics. The difference between the approaches is particularly relevant for long-term patterns, probably because the shortage of data is more critical when capturing long-term phenomena.

As to *Relevance* consensus, the percentage of relevant known rules is greater for *apriori-like*. Once more, it seems that this approach works better at finding well-known associations in this particular dataset. However, the “Relevant and unknown” category seems to be an issue: there are no patterns at all. Prior to consensus, a few cases were annotated by the experts with this category, but it did not appear anymore after agreement. This is unfortunate, since we considered it a very interesting tag. We hypothesize that this happened because the experts found it difficult to annotate cases as relevant, especially those that were frequent but did not conform to conventional patterns.

Table 6 shows the relationship between correctness and relevance. As expected, most of the cases are either totally correct and relevant-known or incorrect and non-relevant.

Tables 7 and 8, present, for each approach, an example of a correct and relevant rule per time gap, as agreed by the three experts. As mentioned above, the average time gap value is expressed in days<sup>3</sup>. For the sake of clarity, healthcare instances have been represented using the textual descriptions of their codes.

As shown in Table 7, the first two patterns are examples of combinations of typical short-term symptoms associated with respiratory disorders, especially at certain times of the year. The third pattern is an example of

**Table 6** Correctness versus relevance percentages

Correctness	Relevance	
	Not Relevant	Relevant
<i>apriori-like</i>		
Incorrect	30	3
Correct	21	39
Par-Temporal	0	7
Par-Clinical	0	0
Par-Both	0	0
<i>patient-oriented</i>		
Incorrect	44.5	1
Correct	0	33
Par-Temporal	4.5	6
Par-Clinical	4.5	4.5
Par-Both	2	0

a very general symptom combined with a body part that provides specific information about the location of the symptom, in this case indicating a chronic osteoarticular pain such as arthrosis. The last pattern is an example of the long-term development of multi-pathology patients with severe chronic cardiac and respiratory pathologies.

Regarding Table 8, the first pattern is an example of a short-term respiratory infection context, which includes fever and is treated with paracetamol. The second pattern shows a combination of symptoms from patients with both respiratory and cardiac conditions. The third pattern could be an example of a possible mid-term side effect, where a respiratory infection treated with paracetamol results in symptoms in the skin and skin appendages. Finally, the last pattern is an example of risk factors typically associated with certain diseases in the long term, in this case possibly patients suffering from paralysis due to HCV or degenerative neurological diseases.

Although some of the patterns in both tables may seem obvious, we consider that they are still interesting as they not only confirm known risk factors, but also provide specific information about the time gap between antecedents and consequent. As mentioned above, the availability of more EHRs should allow our methodology to provide more specific patterns.

Albeit we are not including all the rules of the packs in this article, we would like to illustrate some phenomena with specific examples. For instance, the agreement in Table 3 showed that some results are not easy to interpret. Equation (9) shows an example of a *patient-oriented* pattern that was annotated differently by the experts in both *Correctness* and *Relevancy* aspects. One of the physicians found no relationship between antecedents and consequent, the second one mentioned that

<sup>3</sup> We represent them with equations in which the number appearing above the arrows represents the time constraint parameter (TC), i.e. the average time (in days) between temporal steps.

**Table 7** Apriori-like’s correct and relevant pattern samples, grouped by time constraint type

---

Respiratory symptom/complaint other $\xrightarrow{5}$ Cough $\xrightarrow{6}$ Respiratory symptom/complaint other $\xrightarrow{6}$ Respiratory symptom/complaint other $\xrightarrow{4}$ Throat symptom/complaint
Respiratory symptom/complaint other $\xrightarrow{16}$ Cough $\xrightarrow{16}$ Respiratory symptom/complaint other, Acute nasopharyngitis [common cold]
Entire limb $\xrightarrow{47}$ Musculoskeletal sympt/compl other $\xrightarrow{45}$ Pain general/multiple sites
Entire thorax $\xrightarrow{96}$ Angina pectoris $\xrightarrow{88}$ Pain respiratory system, Entire thorax

---

**Table 8** Patient-oriented’s correct and relevant pattern samples, grouped by time constraint type

---

Shortness of breath/dyspnoea $\xrightarrow{5}$ Fever $\xrightarrow{5}$ Paracetamol
Respiratory symptom/complaint other $\xrightarrow{16}$ Pain general/multiple sites $\xrightarrow{16}$ Cardiovascular sympt/compl other $\xrightarrow{16}$ Shortness of breath/dyspnoea
Respiratory symptom/complaint other $\xrightarrow{43}$ Paracetamol $\xrightarrow{45}$ Skin symptom/complaint other
Psychological symptom/compl other $\xrightarrow{79}$ Pain general/multiple sites $\xrightarrow{87}$ Musculoskeletal sympt/compl other, Neurological symptom/compl. other

---

there was a clear relationship between them, while the third physician stated that both fever and lack of oxygen might cause neurological disturbances (for example convulsions). Eventually, by consensus, it was agreed that the rule was *Partially Correct (clinical facet)* and *Rel-Known* (Relevant and known).

$$\text{Fever} \xrightarrow{5} \text{Shortness of breath/dyspnoea} \xrightarrow{5} \text{Neurological symptom/compl. other} \quad (9)$$

Equation (10) shows another example of an apriori-like pattern with disagreement among the experts: while one saw no relationship, another one thought that limitation in movement might lead to weight gain or, conversely, severe pain might lead to anorexia and weight loss. The third expert saw a debatable relationship. However, the consensus reached was that the pattern was not correct nor relevant. We consider Eqs. (9) and (10) as good examples of the intricacy and subjectivity of the evaluation process.

$$\begin{aligned} \text{Pain general/multiple sites} &\xrightarrow{16} \text{Pain} \\ \text{general/multiple sites} &\xrightarrow{16} \\ \text{Endocrine/met./sympt/compl other} & \end{aligned} \quad (10)$$

Equation (11) shows an example of a patient-oriented pattern that was annotated by the experts as *Rel-Known* and as *Par-Temporal* (Partially correct for the Temporal facet). They argued that the time elapsed among the medical events (in this case, all of the type symptom) was longer than expected, though it is indeed very often found indicating cases of complicated respiratory infections. This is an example of how our system can refute conventional knowledge and indicate that certain events may have a longer time frame to consider for prevention purposes.

$$\begin{aligned} \text{Fever} &\xrightarrow{79} \text{Pain general/multiple sites} \xrightarrow{86} \\ \text{Cough, Respiratory symptom/complaint other} & \xrightarrow{86} \text{Shortness of breath/dyspnoea} \end{aligned} \quad (11)$$

As noted above, some of the patterns were considered novel, i.e. *Relevant Unknown*, by some physicians in the individual ratings, although this assessment did not appear in the final consensus. For example, patient-oriented Eq. (12), which had a high value of the *rules-RespectConsequent* measure (20%), was considered a correct and *Rel-Known* pattern in the consensus. However, in the previous evaluations and during the consensus, two experts commented that they were hesitant about the correctness, but that if this relationship was considered correct, it would be novel to think that 20% of patients who have generalized pain and subsequently some neurological symptom would end up with a diagnosis of malignant neoplasm.

$$\begin{aligned} \text{Pain general/multiple sites} &\xrightarrow{15} \text{Neurological} \\ \text{symptom/compl. other} &\xrightarrow{16} \text{Malignant} \\ &\text{neoplasms} \end{aligned} \quad (12)$$

Finally, Eq. (13) shows another example of a possible evolution of symptoms that were categorized as *Relevant Unknown* by one physician. This patient-oriented pattern may capture a novel set of long-term risk factors typical of a patient with multiple pathologies that evolve in hypertension.

$$\begin{aligned} \text{Cough, Respiratory symptom/complaint other} & \xrightarrow{83} \text{Paracetamol} \xrightarrow{87} \text{Cardiovascular} \\ \text{sympt/compl other, Entire limb} & \xrightarrow{87} \text{Essential} \\ & \text{(primary) hypertension} \end{aligned} \quad (13)$$

In this section, we have just discussed some results. We are aware that our rules show low support, especially for long-term time gaps, as our data set only spans 6 years and these phenomena are more difficult to extract. Obtaining sequential patterns with high support from real-world healthcare data is generally difficult. For this reason, many researchers set a low minimum support. In our case, this situation is exacerbated by the limited dataset we analyzed, which consisted of the EHRs of only 320 patients with multimorbidity. To obtain more robust results, it is essential to have more extensive clinical data, both in terms of longer EHRs and a larger number of patients.

### Generalizability

In this work, we have focused on a very specific type of patient. However, it is certainly possible to use the same algorithms to mine temporal association rules from health records of different patient cohorts or with other medical conditions. In fact, the strength of our method lies in the possibility of extracting patterns from any patient population, provided a few considerations are taken into account:

- **Dataset characteristics:** The input dataset must contain electronic health records with the specific format, coding system and attributes to ensure compatibility with the algorithm. Therefore, depending on the specific patient population or medical conditions, performing additional data preprocessing steps may be required. This could involve handling missing values, normalizing or standardizing data, mapping different coding systems, or adapting the data representation to meet the requirements of the algorithm.
- **Algorithm selection and parameter tuning:** The same algorithm may require different parameter settings when applied to different datasets. It is important to carefully adjust the parameters to optimize the performance and quality of the temporal association rules discovered. Considerations include the minimum support threshold, the minimum confidence threshold, the number of time intervals (and the definition of each interval with a minimum and maximum time gap), and the appropriate approach (apriori or patient-oriented).
- **Medical domain expertise:** When applying our algorithm to different patient cohorts, it is crucial to involve experts in that specific domain. They can provide valuable insights and domain-specific knowledge to guide data preprocessing, algorithm selection and parameter tuning. Experts can also help validate the discovered temporal association rules and assess their clinical relevance and correctness.

By considering these aspects, our algorithm can be adapted to mine temporal association rules from health records of different patient populations or with other medical conditions. However, it is important to note that the results and interpretations may differ due to variations in the data and domain-specific factors. Therefore, thorough analysis, validation, and collaboration with domain experts are key to obtaining meaningful and actionable insights.

### Conclusions

A highly flexible temporal association rule mining algorithm, accompanied by a sound formulation for TARs, has been presented. The algorithm has been designed to be highly parameterizable, not only in terms of support and confidence measures but also in its ability to extract rules with different numbers of temporal steps (each one eventually with different time intervals  $[t_1, t_2]$ ), attributes and levels of concept abstraction.

Using this algorithm, we have successfully extracted temporal association rules that help in the diagnosis and risk prediction of multimorbid patients. To achieve this, certain parameters were fixed, such as the use of codes as attributes, a fixed level of code generalization, and a minimum of 2 temporal steps in the rules. We experimented with different values for other parameters such as support, confidence, and time intervals. We aim to establish a strong basis for the future development of a robust rule extraction tool. Furthermore, due to the adaptability of our approach, it can be easily applied to the extraction of temporal association rules in various other domains.

In addition, our work proposes two different approaches, each with a unique rationale, allowing the choice of the one that best suits the medical dataset and the desired pattern application. The first approach is specifically designed for multimorbid patient datasets (though it might be tested with other patient types). On the other hand, the second approach is designed for conventional datasets and aims to extract rules that are common to patients. In this way, users can choose the approach that suits their specific needs and dataset characteristics.

We implemented a comprehensive evaluation process for the rules obtained. This process involved defining a set of evaluation criteria, agreed by three experts, covering both correctness and relevance. However, the practical evaluation proved to be challenging as it was heavily influenced by each expert's experience and interpretation. To overcome this challenge, we conducted a subsequent consensus step and evaluated Fleiss' Kappa agreement. The results demonstrated the subjectivity of the assessment process.

This evaluation process concluded that the implementations were able to retrieve correct and relevant rules.

The *apriori*-like approach was found to be more accurate in terms of patterns, which may be due to the fact that the dataset used to retrieve the rules consisted of annotations from multimorbid patients with pre-selected characteristics. In some cases, especially in the long term, the rules obtained showed a relationship between pathologies that are highly prevalent. It should be noted that the insight behind such rules may not necessarily reflect a temporal cause-effect relationship, but rather a co-occurrence that signifies the patients' comorbidity. Therefore, it would be beneficial to compare the patterns of multimorbid patients with those of general patients in order to identify multimorbid-specific temporal association rules, similar to what [11] did with diabetic-specific patterns.

It is important to note a final limitation related to the intrinsic code annotations in the dataset, particularly for *Sign or symptom* ICPC-3 codes. The majority of the selected rules contained codes that were too broad (e.g. *Respiratory symptom/complaint other*), resulting in reduced relevance of the rules. This is a recurring problem in our clinical history corpus for two reasons. Firstly, the corpus is relatively small, which means that the most general phenomena tend to be more common. Secondly, primary care physicians are often overwhelmed and tend to assign a generic code because they do not have time to search for the most appropriate code for each case. This lack of time is a chronic problem in all primary care systems, and all physicians complain that EHRs would be more complete and specific if they had more time to enter patient information at each visit. This situation results in "noisy data", such as missing or incomplete data, inconsistent notation, irregular patient visits, and uncertainty about when healthcare events occurred versus when they were recorded. This problem is a drawback of our extraction system. However, an extension of our methodology could eventually provide suggestions and/or additional information to make this process of registering information more efficient. While intelligent diagnosis and treatment machines cannot replace human practitioners, they can assist them in making better clinical decisions.

The limitations of our dataset are related to both its size (in terms of both the number of available patients and the historical data per patient) and its nature (the aforementioned level of specificity of the annotations). Indeed, the size of a dataset can have an impact on the association rules mined from it. An important property that relates the size of the dataset to the association rules is the sparsity of the dataset. Sparsity refers to the proportion of rare itemsets in the dataset. In our case, the data set happens to be both small and sparse. This has meant that there are interesting phenomena that are not captured because the corresponding rules do not meet the minimum support. However, there are very general phenomena that have

been apprehended by the methodology. Therefore we believe that, although larger datasets may lead to the discovery of more diverse and specific association rules, our current findings would still hold with greater reliability beyond our specific dataset. This is a hypothesis, it would be necessary to apply the methodology to an expanded dataset to assess the stability of the rules across different dataset sizes. If the rules obtained from the expanded dataset were consistent or similar to those obtained from the smaller dataset, this would suggest that the rules are robust and not solely influenced by the dataset size.

Despite these limitations, we have successfully extracted accurate and relevant patterns, particularly for the specific group of multimorbid patients. This methodology may be valuable for preventive medicine, as sequential temporal patterns can predict risk factors commonly associated with certain diseases. Therefore, the next step would be to integrate this set of patterns into the existing medical system so that the patterns could be used in the decision-making process. This integration involves representing the patients' medical data in a format suitable for applying the temporal association rules. Then, to assess a patient's risk factor for a disease, the following steps could be taken: (1) select the relevant temporal association rules that are likely to indicate the presence or increased risk of the specific disease of interest (this is indicated in the consequent of each rule), (2) apply the selected temporal association rules to the patient's data and evaluate their applicability, i.e. check whether the antecedents of each rule match any relevant events in the patient's medical history with the rule's temporal constraints, and (3) for each applicable rule, determine the patient's risk factor for the disease. If the rule has high support and confidence, it suggests a potential risk factor for the disease. The strength of the association and the temporal patterns captured by the rules can provide insight into the patient's likelihood of developing the disease. Moreover, the average time gap provides crucial information about the term of these risk factors, which is vital in primary care.

At this stage, the only assessment of the quality of the extracted patterns we can provide is that of the experts. The patterns extracted by our algorithm align with clinical knowledge and can be easily understood by healthcare professionals. In our opinion, this demonstrates the potential of such patterns not only to confirm previously known associations but also to discover new relationships between factors and to better understand hypothesized relationships between such factors. We emphasize the importance of such transparent and explainable models in gaining the trust and acceptance of clinicians, enabling them to make informed decisions based on these patterns, thus empowering personalized preventive measures.

**Abbreviations**

EHR	Electronic health record
KDD	Knowledge discovery in databases
TAR	Temporal association rule
CAP	Primary Care Center (in Catalan)
ATC	Anatomical therapeutic chemical
ICD-10	International classification of diseases 10th revision
ICPC-3	International standard for Primary Care 3rd revision
SNOMED	Systematized Nomenclature of Medicine

**Supplementary Information**

The online version contains supplementary material available at <https://doi.org/10.1186/s12911-023-02287-0>.

**Additional file 1.**

**Acknowledgements**

The authors would like to thank IDIAP JGOL's experts (David Lacasta, Víctor López, Luis Rodríguez, Pere Torán and Concepció Violán) for their evaluation, as well as Dr. Pino Ageno for her valuable comments.

**Authors' contributions**

A.G. and N.C. conceived and designed the study. A.G. and N.C. conducted the literature search. M.P. conducted data collection, data analysis and pre-processing. A.G. and N.C. organized the evaluation process. A.G. and N.C. drafted the paper. A.G., N.C., and M.P. contributed to result interpretation and critically revised the article. All authors approved the final manuscript.

**Authors' informations**

Marcel Pons is a former student of the Facultat d'Informàtica de Barcelona of the Universitat Politècnica de Catalunya, who has participated in this study as his Master Thesis work.

**Funding**

This research was supported by the Spanish Ministry of Science and Innovation, through the TADIA-MED project (<https://futur.upc.edu/28881334/>) [PID2019-106942RB-C33]. The funding body played no role in the design of the study and collection, analysis, interpretation of data, and in writing the manuscript.

**Availability of data and materials**

The data that support the findings of this study are available from Foundation University Institute for Primary Health Care Research Jordi Gol i Gurina (IDIAP JGOL), but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors (AG and NC) upon reasonable request and with permission from IDIAP JGOL.

**Declarations**

**Ethics approval and consent to participate**

The authors confirm that all experiments were performed in accordance with relevant guidelines and regulations. The need for ethics approval and informed consent is deemed unnecessary according to national regulations: Ley Orgánica 15/1999, (Personal Data Protection), Ley 41/2002 (Patient autonomy and rights and obligations in clinical research and documentation) and Ley 14/ 2007 (Biomedical research).

**Consent for publication**

Not Applicable.

**Competing interests**

The authors declare no competing interests.

**Author details**

<sup>1</sup>TALP Research Center, Computer Science Department, Universitat Politècnica de Catalunya (UPC), Barcelona, Spain. <sup>2</sup>IDEAI-UPC Research Center, Universitat Politècnica de Catalunya (UPC), Barcelona, Spain. <sup>3</sup>Facultat d'Informàtica de Barcelona, Universitat Politècnica de Catalunya (UPC), Barcelona, Spain.

Received: 8 May 2023 Accepted: 4 September 2023

Published online: 19 September 2023

**References**

1. Yoo I, Alafaireet P, Marinov M, Pena-Hernandez K, Gopidi R, Chang J, et al. Data mining in healthcare and biomedicine: a survey of the literature. *J Med Syst.* 2012;36(4):2431–48. <https://doi.org/10.1007/s10916-011-9710-5>.
2. Adlung L, Cohen Y, Mor U, Elinav E. Machine learning in clinical decision making. *Med.* 2021;2(6):642–65. <https://doi.org/10.1016/j.medj.2021.04.006>.
3. Lam LHT, Do DT, Diep DTN, Nguyet DLN, Truong QD, Tri TT, Thanh HN, Le NQK. Molecular subtype classification of low-grade gliomas using magnetic resonance imaging-based radiomics and machine learning. *NMR Biomed.* 2022;35(11):e4792. <https://doi.org/10.1002/nbm.4792>.
4. Le VH, Kha QH, Minh TNT, Nguyen VH, Le VL, Le NQK. Development and Validation of CT-Based Radiomics Signature for Overall Survival Prediction in Multi-organ Cancer. *J Digit Imaging.* 2023;36(3):911–22. <https://doi.org/10.1007/s10278-023-00778-0>.
5. Mathur G, Pandey A, Goyal S. A comprehensive tool for rapid and accurate prediction of disease using DNA sequence classifier. *J Ambient Intell Humaniz Comput.* 2022;14:1–17. <https://doi.org/10.1007/s12652-022-04099-y>.
6. Guo C, Chen J. Big Data Analytics in Healthcare: Data-Driven Methods for Typical Treatment Pattern Mining. *J Syst Sci Syst Eng.* 2019;28(6):694–714. <https://doi.org/10.1007/s11518-019-5437-5>.
7. Severac F, Sauleau E, Meyer N, Lefèvre H, Nisand G, Jay N. Non-redundant association rules between diseases and medications: an automated method for knowledge base construction. *BMC Med Informa Decis Mak.* 2015;15(29). <https://doi.org/10.1186/s12911-015-0151-9>.
8. Lin W, Orgun MA, Williams GJ. An Overview of Temporal Data Mining. In: *Proceedings of the Australasian Data Mining Workshop.* Sydney: University of Technology; 2002. p. 83–9.
9. Segura-Delgado A, Gacto MJ, Alcalá R, Alcalá-Fdez J. Temporal association rule mining: An overview considering the time variable as an integral or implied component. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery.* 2020;10. <https://doi.org/10.1002/widm.1367>.
10. Sacchi L, Larizza C, Carlo C, Bellazzi R. Data mining with temporal abstractions: learning rules from time series. *Data Min Knowl Discov.* 2007;15(2):217–47.
11. Concaro S, Sacchi L, Cerra C, Fratino P, Bellazzi R. Temporal data mining for the analysis of administrative healthcare data. In: *Proceedings of IDAMAP 2008 Workshop.* Washington. 2008. p. 75–80.
12. Zhan L, Yu F, Zhang H. A fast algorithm for mining temporal association rules based on a new definition. In: *Proceedings of the 2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)* Guilin: IEEE; 2017. p. 1548–1553. <https://doi.org/10.1109/FSKD.2017.8392995>.
13. Shrestha A, Zikos D, Fegaras L. An annotated association mining approach for extracting and visualizing interesting clinical events. *Int J Med Inform.* 2021;148. <https://doi.org/10.1016/j.ijmedinf.2020.104366>.
14. Harms S, Deogun J. Sequential Association Rule Mining with Time Lags. *J Intell Inf Syst.* 2004;22(1):7–22. <https://doi.org/10.1023/A:1025824629047>.
15. Black W. *Temporal Data Mining in Electronic Medical Records from Patients with Acute Coronary Syndrome* [Ph.D. thesis]. University of Washington; 2013.
16. Gharib TF, Nassar H, Taha M, Abraham A. An efficient algorithm for incremental mining of temporal association rules. *Data Knowl Eng.* 2010;69:800–15. <https://doi.org/10.1016/j.datak.2010.03.002>.
17. Agrawal R, Srikant R. Fast Algorithms for Mining Association Rules in Large Databases. In: *Proceedings of the 20th International Conference on Very Large Data Bases (VLDB '94).* San Francisco: Morgan Kaufmann Publishers Inc.; 1994. p. 487–499.
18. Wu X, Kumar V. *The Top Ten Algorithms in Data Mining.* 1st ed. Florida: Chapman & Hall/CRC; 2009.
19. Agrawal R, Srikant R. Mining sequential patterns. In: *Proceedings of the Eleventh International Conference on Data Engineering.* Taipei. 1995. p. 3–14. <https://doi.org/10.1109/ICDE.1995.380415>.

20. Srikant R, Agrawal R. Mining sequential patterns: Generalizations and performance improvements. In: Apers P, Bouzeghoub M, Gardarin G, editors. *Advances in Database Technology, EDBT 1996. Lecture Notes in Computer Science*, vol. 1057. Berlin, Heidelberg: Springer; 1996. p. 1–17.
21. Mooney C, Roddick J. Sequential pattern mining - approaches and algorithms. *ACM Comput Surv.* 2013;45(2):1–39. <https://doi.org/10.1145/2431211.2431218>.
22. Ji Y, Ying H, Tran J, Dews P, Lau S, Massanari R. A functional temporal association mining approach for screening potential drug-drug interactions from electronic patient databases. *Inform Health Soc Care.* 2016;41(4):387–404. <https://doi.org/10.3109/17538157.2015.1064427>.
23. Nguyen D, Luo W, Phung D, Venkatesh S. LTARM: A novel temporal association rule mining method to understand toxicities in a routine cancer treatment. *Knowl-Based Syst.* 2018;161:313–28. <https://doi.org/10.1016/j.knosys.2018.07.031>.
24. Orphanou K, Dagliati A, Sacchi L, Stassopoulou A, Keravnou E, Bellazzi R. Incorporating repeating temporal association rules in Naïve Bayes classifiers for coronary heart disease diagnosis. *J Biomed Inform.* 2018;81:74–82. <https://doi.org/10.1016/j.jbi.2018.03.002>.
25. Guillame-Bert M, Dubrawski AW, Wang D, Hravnak M, Clermont G, Pinsky MR. Learning temporal rules to forecast instability in continuously monitored patients. *J Am Med Inform Assoc.* 2017;24(1):47–53.
26. IDIAP. IDIAP Jordi Gol. 2022. <https://www.idiapjgol.org/index.php/en/>. Accessed 13 Feb 2023.
27. Neo4j. Neo4j - The World's Leading Graph Database. 2012. <http://neo4j.org/>. Accessed 15 Feb 2023.
28. Extended Association Rules Source Code [source code]. <https://doi.org/10.5281/zenodo.8141389/>.
29. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull.* 1971;76(5):378–82.
30. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33(1):159–74.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

