



**HAL**  
open science

## Enhanced visual data mining process for dynamic decision-making

Hela Ltifi, Emna Benmohamed, Christophe Kolski, Mounir Ben Ayed

► **To cite this version:**

Hela Ltifi, Emna Benmohamed, Christophe Kolski, Mounir Ben Ayed. Enhanced visual data mining process for dynamic decision-making. Knowledge-Based Systems, 2016, 112, pp.166-181. 10.1016/j.knosys.2016.09.009 . hal-03280476

**HAL Id: hal-03280476**

**<https://uphf.hal.science/hal-03280476v1>**

Submitted on 21 Dec 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Enhanced visual data mining process for dynamic decision-making

Hela Ltifi<sup>1,5</sup>, Emna Benmohamed<sup>1</sup>, Christophe Kolski<sup>2,3,4</sup>, Mounir Ben Ayed<sup>1,6</sup>

<sup>1</sup> REsearch Groups in Intelligent Machines, University of Sfax, National School of Engineers (ENIS), BP 1173, Sfax, 3038, Tunisia.  
hela.ltifi@ieee.org, mounir.benayed@ieee.org.

<sup>2</sup> Univ Lille Nord de France, F-59000 Lille, France,

<sup>3</sup> UVHC, LAMIH, F-59313 Valenciennes, France,

<sup>4</sup> CNRS, UMR 8201, F-59313 Valenciennes, France

<sup>5</sup> Computer Science and Mathematics Department, Faculty of sciences and techniques of Sidi Bouzid, University of Kairouan, Tunisia

<sup>6</sup> Computer Science and Communication Department, Faculty of Sciences of Sfax, Route Sokra Km 3.5- BP 1171-3000 Sfax, Tunisia

---

## Abstract

Data mining has great potential in extracting useful knowledge from large amount of temporal data for dynamic decision-making. Moreover, integrating visualization in data mining, known as visual data mining, allows combining the human ability of exploration with the analytical processing capacity of computers for effective problem solving. To design and develop visual data mining tools, an appropriate process must be followed. In this context, the goal of this paper is to enhance existing visualization processes by adapting it under the temporal dimension of data, the data mining tasks and the cognitive control aspects. The proposed process aims to model the visual data mining methods for supporting the dynamic decision-making. We illustrate the steps of our proposed process by considering the design of the visualization of the temporal association rules technique. This technique was developed to assist physicians to fight against nosocomial infections in the intensive care unit. Actually, an evaluation study *in Situ* was performed to assess the automatic prediction results as well as the visual representations. At the end, the test of the efficiency of our process using utility and usability evaluation shows satisfactory.

*Keywords:* dynamic decision-making; visualization; data mining; cognitive modelling; association rules

---

## 1. Introduction

In today's real-time applications, temporal data is produced at increasing rates. In this context, dynamic decision-making, visualization and data mining are important trends and acquire a more and more significant place in research works [5] [9] [28] [29] [45]. As a matter of fact, visualization and data mining integration in the dynamic Decision Support Systems (DSS) reduces subjectivity and provides a new interesting visual knowledge for decision-makers [49] [50].

Visualization and data mining technologies must closely be associated to each other providing the visual data-mining field [32] [45] [50]. In fact, the contribution of data mining is in the useful patterns identification to provide support for dynamic predictions to be integrated into the decision-making process for decision planning [51]. While the data mining methods offer expertise, temporal visualization assists data miners in understanding the different levels of hidden information in large-scale temporal data [28] [40] [71].

To design and develop visualization techniques, we must follow different steps of the visualization process with regards to Human computer interaction. It is the process of generating views and producing appropriate

interaction techniques for a specific data set. Indeed, the research on visualization should not stop at designing and generating representations but should also study how visualization affects decision-maker thinking, comparing and reasoning [44], by referring to a cognitive model, such as those proposed by J. Rasmussen [55] or one of its variants [27] [36]. The idea is to support visual data mining based decision-making in dynamic situations by making use of temporal and cognitive modelling in the visual data mining process.

Among the application areas, the medical decision-making is a very important field of research [3] [56], and we have noticed that numerous useful visual intelligent techniques have been designed in this area (see for instance [2] [4] [35] [44]). In this paper, we focus on the dynamic decision-making for the fight against Nosocomial Infections (NI) in the hospital intensive care units where a large volume of temporal data is daily generated. Such dynamic data sets have important potential in describing changes in the patterns of appearance of nosocomial infections. One of the popular data mining and knowledge discovery strategies is the association rules technique that represents knowledge in the form of relations between variables.

This paper is organized as follows. In Section 2, we briefly describe the theoretical background of our research including visual data mining, decision-making and visualization processes. In section 3, we show how the temporal visualization process can be adapted under the cognitive considerations (in relation with the user perception abilities) and the visual data mining tasks. Our adapted visualization process is illustrated in section 4 by considering the modelling of the visual temporal association rules that we have developed for the fight against nosocomial infections in the intensive care units. Then section 5 presents the results of the evaluation study of our application. And finally, section 6 outlines our conclusion and several research perspectives.

## 2. Theoretical Background

In this section, we present the theoretical background to provide the rationale for conducting our research. In fact, it consists of reviewing the concepts of the visual data mining for the decision-making, the visualization processes as well as the existing cognitive approaches.

### 2.1. Visual data mining for dynamic decision making

Integrating visual data mining methods in the real-time decision-making is becoming essential for handling complex and temporal data. In fact, organizations daily collect and generate huge temporal data amounts. Applying data mining techniques [25] allows automatizing the processes of extracting interesting temporal knowledge and regularities from electronic records and providing indications to the decision-making [6] [12] [20] [23].

Using temporal visualization methods allows: (1) involving graphical data presentation and interpretation, (2) providing matching results, (3) helping in temporal patterns identification, (4) indicating changes in massive data pools during time, and (5) enhancing cognitive perception [69]. The main aspects of temporal data visualization are the interactivity and the dynamics of visual representations.

Therefore, visual data mining consists of visual analytics of data to automatically and visually discovering connections and hidden relationships from apparently unrelated information. It also enables decision-makers to visually predict trends in the temporal data and their behaviours. Visual data mining techniques enhance dynamic decisions by presenting automatic results in the form of visual and temporal interpretations to understand complex scores (cf. Fig. 1).

Our research area is characterized by its interdisciplinary nature including visual data mining methods, which are applied to temporal data for automatizing analysis. It uses data mining techniques for patterns extraction from temporal data, which can be interactively visualized to evaluate and refine data parameters.

Combining automatic and visual methods characterizes the visual data mining process for the dynamic decision-making. Visual interpretations of data and/or patterns allow steering the model building in the automatic data mining. These visually validated patterns must be integrated as new knowledge for decision-making. Such integration aims to develop alternative solutions for problem solving, by allowing the fusion of extracted knowledge and knowledge of experts (decision-makers). Feedback is possible if the decision maker decides that already integrated knowledge should be considered or refined by re-running the visual data-mining algorithm.

The developed solutions are analyzed in order to consider a best alternative that seems to be the most rational for solving the decision problem. Such analysis must take into account time and decision constraints. The chosen alternative must be communicated to the decision-maker. Then the decision must be converted into an effective action. Without such conversion, the decision will remain just a declaration of good intention (cf. Fig. 1).

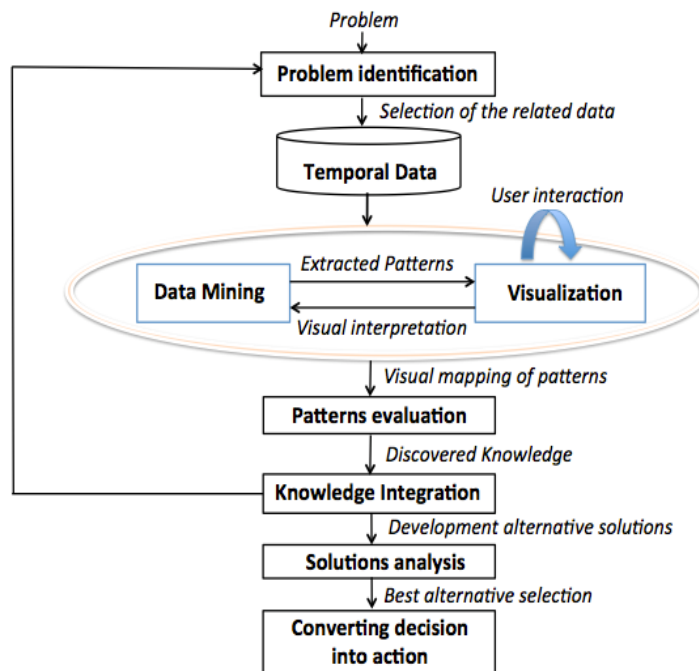


Fig. 1. Visual data mining for dynamic decision making

Visual data mining for dynamic decision-making research can be seen as an integrating discipline [60], soliciting related research fields must contribute with their existing methods and models. Therefore, the requirements of visual data mining introduce new dependencies between the concerned fields (cf. Fig. 2).

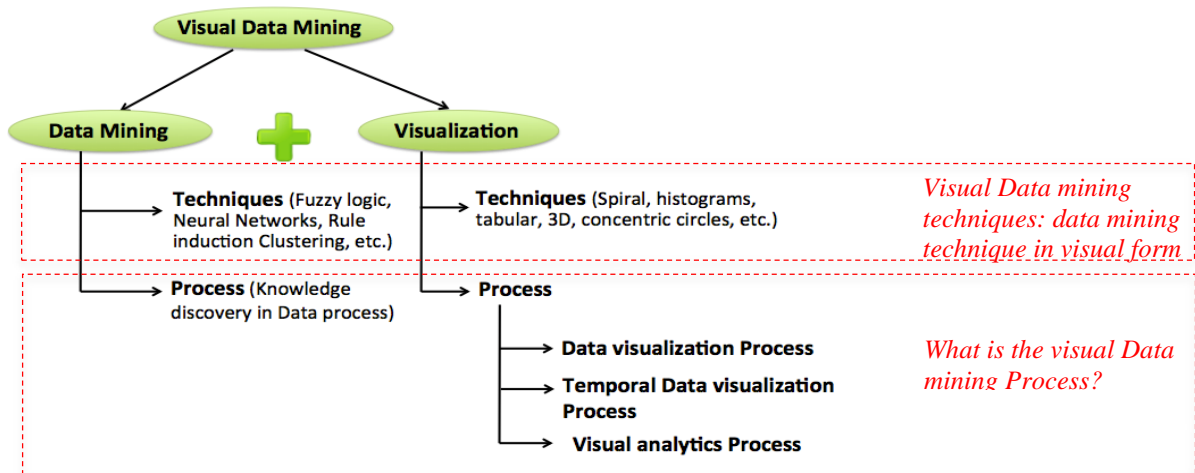


Fig. 2. Research investigation about visual data mining process

To respond to such research investigation (cf. Fig. 2), we begin by examining a set of existing visualization processes proposed in the literature.

## 2.2. Visualization processes

### 2.2.1. Data visualization process

According to Chi [14] a data visualization process can be recognized within every visualization technique. It passes through three transformation phases: (1) Filtering to transform raw data into analytical abstraction using appropriate analytical computations (interpolation, filter techniques, reduction, etc.), (2) Mapping to transform resulting analytical abstraction into a visual abstraction (geometrical objects), and (3) Rendering to transform geometrical abstraction into visual form projected onto display (cf. Fig. 3).

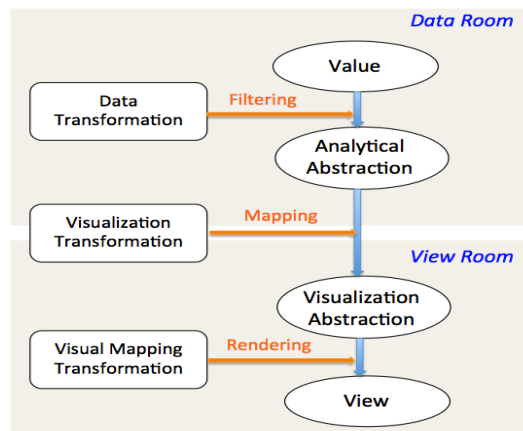


Fig. 3. Visualization process adapted from [14]

Our research concerns visualization techniques for visually analysing and handling temporal data. Chi's visualization process doesn't take into account the temporal aspects (temporal dimension, values, time, etc.). It is therefore difficult to envision and to explore temporal visualization techniques. Addressing this problem, Daassi et al. [16] propose to adapt the Chi's visualization process, as explained below.

### 2.2.2. Temporal data visualization process

Daassi et al. [16] proposal is based on the Chi's process. These authors distinguish two dimensions of temporal data: the structural dimension characterized by a data type and the temporal dimension. Daassi's visualization process structures each dimension process into four steps (cf. Fig. 4):

- (1) Data: represents the temporal or structural values to be visualized. For example, as illustrated in Fig. 4, the data could be an antibiotic to daily take.
- (2) Point of view on data: defined by the transformation of the temporal or structural values to the appropriate representations. For instance, in Fig. 4, the point of view on time is multi-granular (periodic and linear) while point of view on the structural value is quantitative.
- (3) Data space: defines the displayable space of the temporal or structural values. In the visualization process shown in Fig. 4, the antibiotic quantities (the point of view on the data) are ordered (periodically). This transformation fixes the characteristics of the graphical space in which the data are visualized in terms of the number of its dimensions.
- (4) Point of view on the data space: refers to the perceptible translation of the temporal or structural dimension by defining the representations according to the human perception of data. In the example of Fig. 4, concentric circles representation combined with a timeline is defined for presenting the antibiotics.

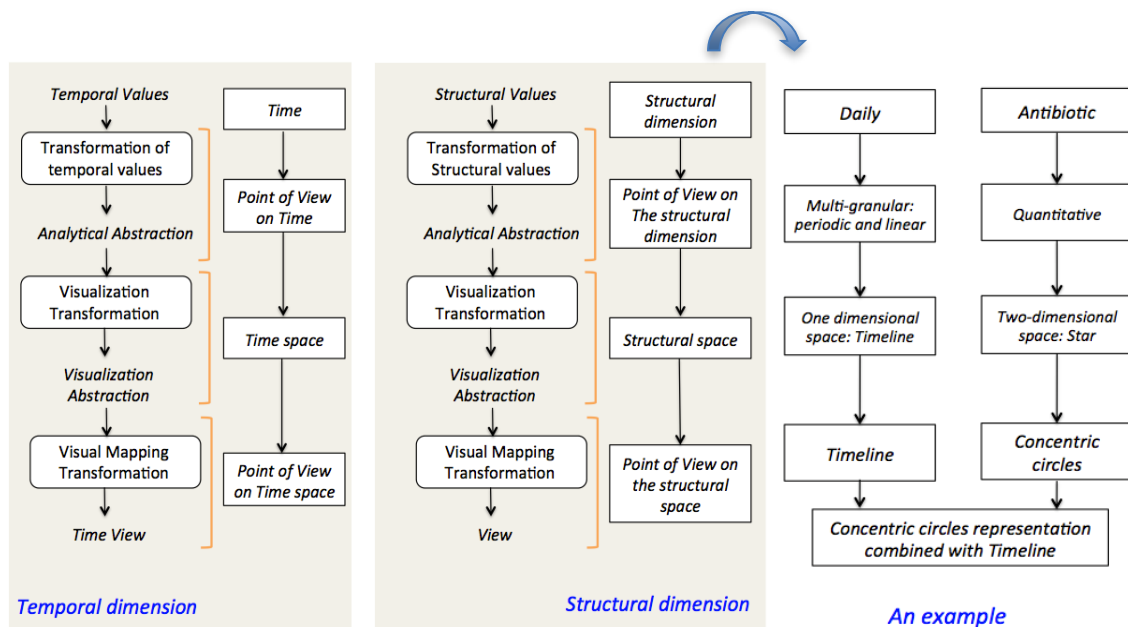


Fig. 4. Visualization process adapted from [16]

As a matter of fact, the choice of the appropriate visualization technique is influenced by the point of view on time. This step identifies the characteristics of the temporal values to be represented (periodicity and

dimensionality). However, this process does not consider the data mining tasks for decision-making. Consequently, we survey a proposed visual analysis process in the next section.

### 2.2.3. Visual analytics process

Visual Analytics is the field of using visual information representations for analytical, knowledge discovery and decision-making purposes. Researchers also use this term to refer to the Information Visualization and Visual Data Mining [33] [64] [65] [66] [72] [73].

The visual analytics process [72] allows producing visual representations from raw data for the decision-maker (user of the system) in order to acquire and interact with knowledge. This process consists of a series of steps and begins with initial analysis including data analysis or mining. It then enters a loop where the user can gain knowledge on the data under study. Besides, it drives the system toward more suitable analytical methods. By interacting on the visualization, the user gain a better understanding of the representation itself ordering different views assisting him/her to go beyond the visual and ultimately confirm results extracted from previous iterations (cf. Fig. 5).

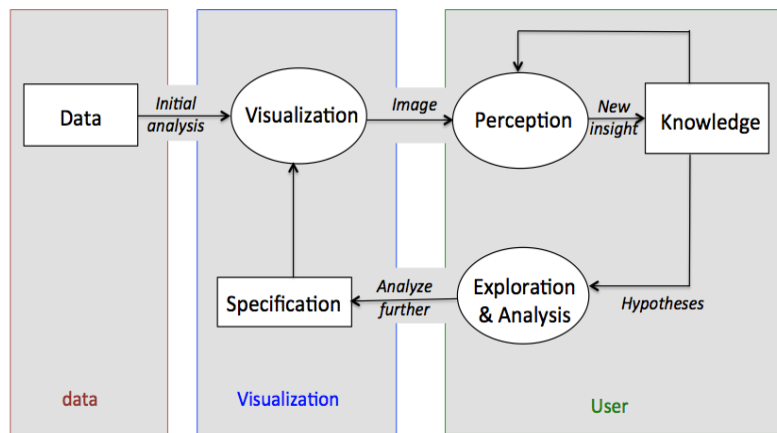


Fig.5. Visualization pipeline for visual analytics adapted from [72]

The visual analytics process visible in Fig. 5 considers analysis and visual aspects (such as perception and insight) and provides description of how users interact with visualization for analytical purposes. However, the cognitive characteristics of the user behaviour are not clearly presented in the process steps as well as the specific data mining tasks. A brief review on cognitive approaches in literature is presented in the following section. And these approaches address concepts ranging from perception to problem solving and decision-making.

### 2.3. Cognitive approaches for dynamic decision-making

Cognitive approaches for dynamic decision-making deal with the data analysis or mining [34] [70]. The identification of the human cognitive processes is a critical issue since it argues that human (decision-maker) behaviour drives from cognitive processes organizing a series of data handling as a reaction to a given stimulus.

In cognitive approaches, cognitive descriptions of behaviour are made based on the ways in which decision-makers organize, control and treat information significant to specific actions. Such descriptions can

be designed using cognitive models. The latter aim to represent the decision-makers by addressing and formulating cognitive activities such as understanding, intentions, processing and knowledge.

Decision-makers are integrated as part of the dynamic situations in decision-making activity, involving their own dynamics and reacting to human operators' actions [27]. In these dynamic situations, operators include their plans, resources representations and potential risks. Salas et al. [58] and Cannon-Bowers et al. [13] works underline the importance of coordination among team members.

Dearden et al. [15] propose dynamic or adaptive task allocation to the machine by assuming that decision-makers consider the task to be allocated as unitary and associated with an intention to be covered [48]. Jungerman et al. [30] specify a cognitive model for understanding the dynamics of cooperative decision-making situations examining, in particular, the acceptance or the rejection of an expert advice.

In fact, cognitive modelling is related to trust. Therefore, a sophisticated understanding of trust is necessary to understand how it influences decision-making in different ways. Trust modelling allows building strong confidence in every step of the decision-making process. Thus, as trust is considered as the mechanism of reducing social complexity and resolving uncertainty [37] [38], several trust models have been proposed in various areas. For example, [10] [67] suggest trust assessment model in an open dynamic, untrusted and also uncertain environments. Yu et al. [74] present more details on the trust models available in the literature.

In the context of data mining based decision-making, attributing the appropriate quantity of trust to extracted models is necessary for integrating them to make decision. That's why, appropriate quantitative trust measures must finally reflect the probability that the model predictions correspond to future test targets. Nevertheless, due to the exploratory nature of most data mining tasks, fully articulating trust probabilistic factors generally seems intractable. Hence, instead of focusing on trying to measure trust in probabilistic quantity, it is naturally useful to visualize many dimensions of the key factors that contribute to trust in those extracted models. In addition, visualizing the limits of a scientific model is of paramount importance. Indeed, we can better see the overall aim to "*visualize trust*" than to understand the limitations of the model [63].

Simoff et al. [60] propose two strategies of assessing human cognition in visual data mining: (1) the guided cognition to simulate the human cognitive mechanisms for discovering the visual forms, interacting with them and conducting meaningful interpretations; and (2) the validated cognition to confirm the results of the human perception of the visualized forms.

Hence, it is important to consider the concepts of: (1) dynamic situation, cooperative modelling, tasks allocation and trust assessment in dynamic decision-making, and (2) supporting human cognition in visualization and data mining. This kind of approaches has limited applicability because our context is more complex with the high temporal constraints and the heterogeneous theoretical frameworks including decision-making, visualization and data mining tasks. In this context comes our contribution.

#### **Motivation:**

This work is motivated by the facts that:

1. Visualizations are subjective, and non-expert decision-makers are frequently left to use visualization methods without the ability to comprehend their applicability and limitations.
2. The design of consistent visualizations must involve the tasks of the visual data mining.

To overcome these challenges, we propose to integrate cognitive and temporal aspects in visual analytics process (described in section 2.2.3) in order to model the decision-makers behaviours in dynamic situations and to structure the design space of the used visualization techniques. We aim, in particular, to take advantage of the positive elements in the fields of visual data mining and cognitive modelling to develop Decision Support Systems in dynamic situations.



### 3. Cognitive and temporal enhancement of the visual data mining process

The goal of the visual data mining methods in the dynamic situations is to allow the decision-makers to interactively search temporal information, detect important relations and identify interesting models to be used in the dynamic decision-making. The visualization supports this process by involving decision-makers to exploit their human capacity for perceiving, abstracting and understanding available complex data. This context involves interdisciplinary fields as shown in Fig. 6.

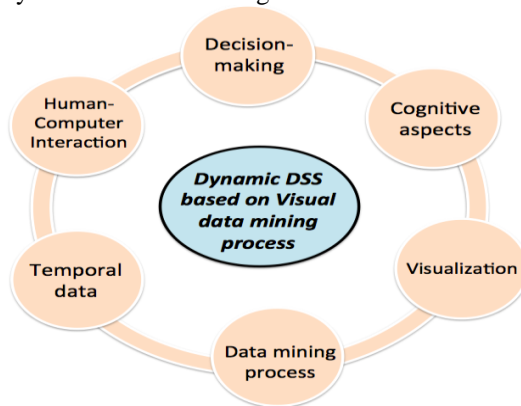


Fig. 6. Involved interdisciplinary fields

#### 3.1. Contribution presentation

In this paper we aim to propose a specific process for the design and development of visual data mining tools, which produce useful knowledge for dynamic decision-making. In such process, we aim to integrate data mining characteristics and human aspects. And it must define and relate relevant concepts to produce knowledge from temporal data to decision and vice versa. This process incorporates these concepts into an iterative framework and demonstrates possible human machine couplings, which are fundamental in the visual data mining and interactive DSS. It must not be limited to visual analysis but also be relevant in decision-making task.

Our proposition aims to give a basis for more detailed compositions of temporal information, visual data mining and DSS theories. Although it is not our intention to cover all the unique details of these theories, we plan to provide a framework that meets these theories as presented in Fig. 6. To do so, instead of proposing a new process, we think adapting existing process under other theories could be a practical solution. In this axis, our theoretical proposal consists of adapting the visual analytics process of [72] under the temporal, data mining and cognitive considerations. Such adapted process suggests three main steps to be followed by the decision-makers for creating enhanced analytic visualizations.

#### 3.2. Step 1: temporal data manipulation

The proposed process begins with a set of defined goals and objectives of the real-time decision-making problem. It is important that related temporal data sources be selected. While raw data can contain errors and missing values [19], it also seems critical to prepare them in order to construct the temporal data set that will be used for exploration. In fact, applying data mining on data, which has not been carefully assessed for

complex dynamic problems, can generate highly misleading patterns. Therefore, prediction results depend on the quality of the prepared temporal data.

“A temporal data denotes the evolution of an object characteristic over a period of time” [16]. It is a sequence of  $\langle t, v \rangle$  pairs where  $t$  denotes the time value (i.e. day, hour or minute) and  $v$  its structural value (i.e. a number, a colour, an object, etc.). A temporal data is defined as the combination of temporal and structural dimensions (cf. section 2.2.3).

Temporal data transformation aims to let the user (i.e. decision-maker) understand what data to manipulate. It is actually an analytical transformation and it consists of extracting from the temporal values the time representation (granularity, periodicity and dimensionality) and from the structural values the data representation.

Data mining techniques will be applied to the prepared and transformed temporal data to extract models that might help decision-makers to interactively evaluate and make appropriate alternatives. Added to that, in dynamic DSS, the choice of the data mining technique must consider the dynamic aspect of the decisions as explained in the section 2.1 [43] [47].

### 3.3. Step 2: temporal visualization

The data mining technique(s) involve(s) extracting useful models from the prepared and transformed temporal data, which is not already known for the user. These output models take the form of probabilities, decision trees, rule induction, regression models or neural networks [25]. They are characterized by their accuracy and understandability [41] especially with their temporal aspect. It is then important to translate them into a visual solution.

As there are several ways to graphically represent a temporal model, the used visualizations must be selected to maximize the value for the user. The design of such visualizations involves the understanding of the user's needs. Since the viewer (i.e. decision maker) is not an expert in data modelling, it seems crucial to translate the extracted model into a more natural image for him/her. For this purpose we propose to design the time space and the structural space of the model by:

1. Transforming the time values onto displayable time space (such as a timeline\* or a perspective wall†; for different types of information presentation, see [1] [2] [16] [45]).
2. Transforming the analytical representation of the structural data onto values prepared to be displayed. This transformation identifies the features of the graphical space in which the data are represented in terms of the dimensions' number.

The result of this step is a model visual mapping composed by a perceptible rendering of its temporal dimension combined with a perceptible rendering of its structural dimension. Such combination must fit together and must inspire confidence as the visualization is traversed. It must include the specification that includes the ability to navigate data values over time and to explore in some optimal way the interaction modes such as zooming, filtering, labelling (cf. Fig. 7).

---

\* Timeline: It allows displaying a list of events in chronological order [26].

† Perspective wall: This 3D technique allows data to be presented chronologically on several panels [46].

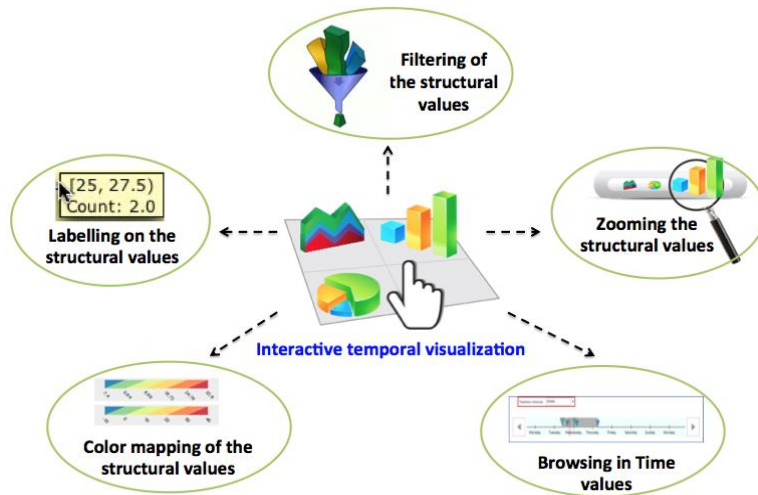


Fig. 7. Interaction modes of interactive temporal visualizations

### 3.4. Step 3: Discovered Knowledge management

Once the visualization is provided, the decision-maker catches sight of the image. Such human perception is affected by the initial system objectives but also by the user abilities, domain experience and performed trainings [55]. These are cognitive aspects, which the decision-makers can depend on to cognitively perceive the data-mining model.

Behind the model perception, the decision-maker must understand it. A model comprehension consists of utilizing the model to comprehend what is actually going on. It must involve the context. Visualizing the output model allows each decision-maker discussing and describing the logic behind it with the other decision-makers. The comprehension of the visual model logic is part of building the decision-makers' trust in the results. If the decision-makers can understand what has been discovered in the context of their decision issues, they would trust it and make it into use. Ultimately, knowledge is the direct result of comprehending and trusting the temporal model, which is with the use of the problem solving.

The knowledge management step depicts the knowledge as feeding interactive exploration since the decision-maker interacts within the temporal visualization using the specification consisting of a set of manipulation techniques.

As visible in Fig. 8, we have proposed two additional arrows to the model initially proposed by [72] (visible in Fig. 5):

1. From Perception to Exploration: indicate the critical role that perception plays in active interactive exploration and analysis.
2. From Exploration to Knowledge: show how interactive exploration feeds the knowledge reasoning. As the human decision-maker explores to learn, that learning directs the attention for more exploration.

The manipulation of the visualization dynamically allows the decision-maker to get a feel for the test whether something really counter-predictive happens. The *Exploration-visualization-perception* process continues until the decision-maker could see the actual knowledge behind the model. Finally, knowledge will be integrated within decision support.

Mathematically, [72] expressed Knowledge as:

$$K(t) = K_0 + \int_0^t P(I, K, t) dt \quad (1)$$

The image  $I$  is perceived by the user, with an increase in knowledge  $K$  over time. The current knowledge  $K(t)$  follows from integration over time, where  $K_0$  is the initial knowledge. The equation (1) means that the amount of gained knowledge depends on the image, the current user knowledge, and the user perception  $P$ . For example, a physician can extract more information from a medical image than a non-expert person. Nevertheless, when such initial knowledge is available, the added knowledge shown in the image can be low.

Where Knowledge ( $K(t)$ ) is the extension of the held knowledge ( $K_0$ ) through the perception of the Image ( $I$ ). This equation does not consider the idea of the cognitive reasoning whereas the model perception, comprehension and trust form a fundamental step of the process. Indeed, our adaptation would define the knowledge creation over time as indicated in the following equation:

$$K(t) = K_0 + \int_0^t E(P, C, T, K, t) dt \quad (2)$$

Where Knowledge ( $K(t)$ ) is the extension of the held knowledge ( $K_0$ ) through the integrated perception ( $P$ ), comprehension ( $C$ ) and trust ( $T$ ) during the cognitive processes of Exploration.

An important aspect is the interactive exploration (represented by  $E$  in the equation (2)). The user could decide to adapt the visualization specification, based on his/her current knowledge, in order to explore the image perception further based on his/her comprehension and trust on the perceived data. Hence the current knowledge  $K(t)$  follows from integration over time where  $K_0$  is the initial knowledge.

For example, if a physician is responsible for monitoring a patient, understanding his/her temporal vital parameters, graphically represented (the image), is critical. And the decision about what real-time action to take is a direct result of understanding the image perception. There is no automated way to do this since it is all in the physician's head. If the physician can understand what has been perceived in the medical context, he/she will explore building trust for the creation of knowledge and put it into use.

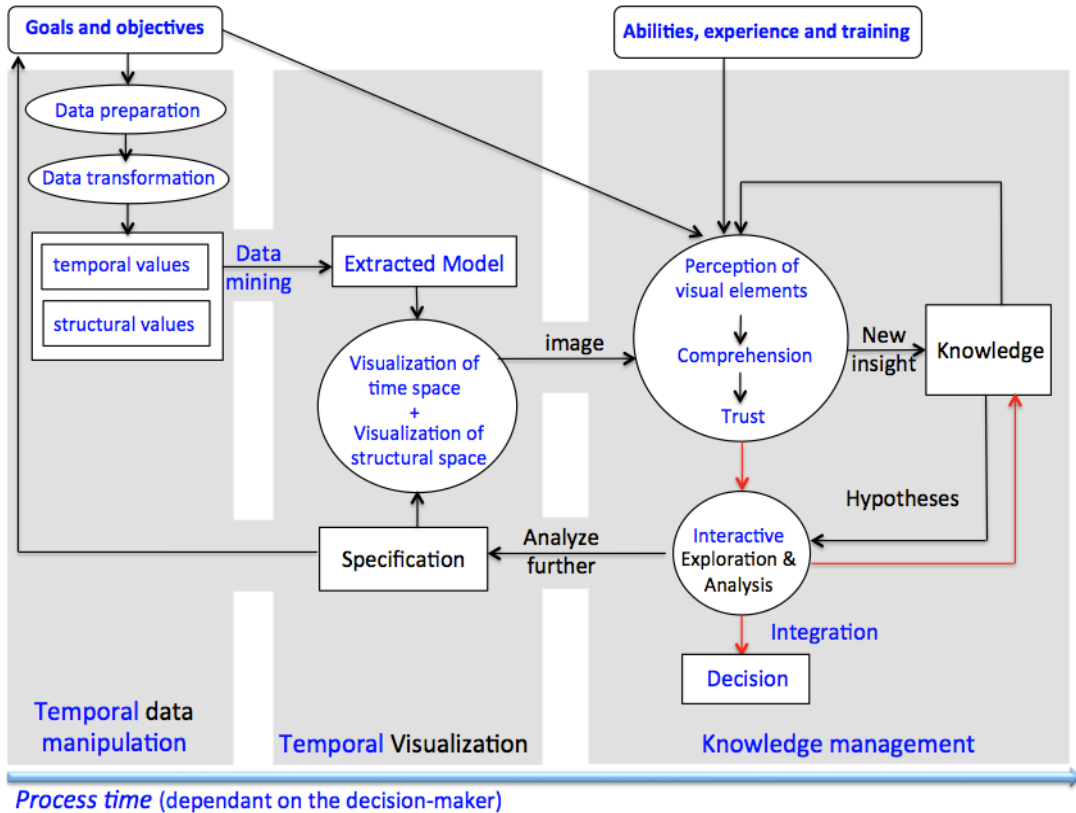


Fig. 8. Visual data mining process

The proposed process is applied to design a visual and temporal association rules tool used to fight against nosocomial infections in intensive care units.

#### 4. Case study: Design of a temporal association rules tool for the fight against nosocomial infections

In this section, we begin by presenting the problem of nosocomial infections and the importance of developing a visual temporal association rules to prevent them. Then, we will present the application of the proposed process to develop the selected visual data mining technique.

##### 4.1. The problem of nosocomial infections

Nosocomial Infections are considered as a real challenge for health authorities [18] [61] [62]. Moreover, an infection is considered as nosocomial if it is acquired in hospital, that appears 48 hours or more after the hospital admission. In fact, the frequency and the gravity of the nosocomial infections are important indicators for assessing the health system development status in a given country.

Our application context considers this problem in the Habib Bourguiba Hospital of Sfax, Tunisia, where the nosocomial infections prevalence was estimated at 9.8% in the study of [21]. According to these authors,

the prevalence was in the Intensive Care Unit (42.1%), then the dermatology (18.5%), rheumatology (17.9%) and nephrology (16.7%) services.

The daily nosocomial infection surveillance is essential to constantly promote prevention and to control patient's state during their hospitalization [17] [53]. To do so, we propose to design and develop a visual data mining based dynamic decision-making tool for the fight against nosocomial infections in the hospital intensive care unit.

#### 4.2. The visual association rules

The association rules mining can be considered as well used data mining technique for decision-making [31] [59]. It is the process of finding interesting and unexpected rules of the form of  $A \Rightarrow B$  from temporal data. It allows the prediction of the occurrence of an item based on the occurrence of the others. The quality of a rule can be assessed by the two criteria: (1) the confidence to measure the certainty of the rules, and (2) the support to measure the usefulness of the discovered rules.

In this paper, we are interested particularly in the graphical representation of the temporal associations for the relevant knowledge extraction from a large mass of temporal data. Although the textual representation (natural language) can be easily understood, the cognitive effort exerted to interpret a large number of rules remains high; therefore we propose to explore this knowledge in an interactive visual space. To do so, we apply the proposed cognitive process to develop these dynamic association rules for the fight against nosocomial infections.

#### 4.3. The application of the proposed visual data mining process

The application of the visual data-mining pipeline for the design of the visual temporal association rules follows the steps visible in the Table 1.

Table 1. Application of the visual data mining process

| Step                       | Decision-maker requirements                                       | Description   |
|----------------------------|---|---|
| Temporal data manipulation | Daily analyzing the patient's temporal data for the NI prevention | <p>The prescribed goal of this application is to develop a visual and temporal association rule mining technique for the decision support of the nosocomial infections surveillance in the ICU. We plan to daily analyze the patient's temporal data. Such analysis must be interactive and visual to gain a better understanding of the patient state evolution.</p> <p>The decision-making system is developed in collaboration with the ICU team. This ICU produces daily large amounts of temporal data. For each patient, a set of control measures is taken daily during his/her hospitalization: the antibiotic catch, the performed medical acts, the infectious examinations and the biological parameters. These data are daily collected from a variety of sources. For this reason, we found some partially missing data because of temporary unavailability of ICU equipment. To solve this problem, we have applied a first prototype for temporal data preparation based on the techniques of clustering and soft computing [68].</p> <p>After data pre-processing, we looked more closely at their transformation. The time unit of the temporal dimension is the Day. The structural dimension of each kind of data is dependent on the data type (1, 2 or 3D, quantitative or qualitative and linear or periodic)</p> |
| Data mining                | Daily calculating NI probabilities                                | We have coded an association rules mining algorithm taking into account the temporal nature of data. We attempt to improve the classical association rules by exploring the   |

|  |  |
|--|--|
|  | <p>time factor of data. Examples of useful rules produced by our algorithm are of the form: <i>if act X Then NI probability in T</i>:</p> <ul style="list-style-type: none"> <li>• If <i>Artificial Ventilation</i> on 07/10/2014 and <i>Tracheal intubation</i> on 09/10/2014 then <i>NI probability</i> in 28 days.</li> <li>• If <i>Urinary Probe</i> and <i>Intravenous injection</i> on 07/01/2015 then <i>NI probability</i> in 7 days.</li> </ul>   |
| <p>Temporal visualization</p> <p>Interactively Visualizing temporal data and knowledge (in terms of probabilities)</p>   | <p>In this step, we begin with the definition of the displayable space of the time values observed at the unit “Day”. The graphical representation of the generated rules is designed within a three-dimensional space. We have mapped into the X-axis of 2D matrix the days of the hospitalization stay, into the Y-axis the items (acts) and the conclusion (NI), then we have represented the metrics (support and confidence) using histograms.</p> <p>The specifications of the temporal visualization include the use of the following developed interaction features:</p> <ul style="list-style-type: none"> <li>• Inspecting the rules by selecting the required region in the visualization space.</li> <li>• Zooming in or out a selected region.</li> <li>• Filtering the acts to display only those that are dominant.</li> <li>• Sorting the rules according to the support or confidence</li> </ul>  |
| <p>Image</p> <p><i>Interactive temporal data visualization</i></p>   | <p>The visualization of the temporal association rules is presented in the Fig. 9. As visible in the figure, we visualize association rules in several main dimensions: Items/conclusion, Time, Support and Confidence. Taking an act at a time <math>t</math> is expressed by the intersection of the X-axis and the Y-axis (2D matrix). Each cell of the matrix represents a rule, and we have used the space operated by the box to represent it. Metrics (Support and Confidence) are shown in the bottom of the array, and their sizes are proportional to their values.</p> <p>The rules are distinguished by different colours, i.e. the premises and the conclusion of the same rule are made by same colour. For example, the green colour in the 3D Histogram representation (in the Fig. 9) displays the following rule: If Tracheal Intubation (12/07), Drainage LC (14/07, 15/07, 16/07), Tracheotomy (19/07), conclusion=Infected (82%) (20/07) with support = 15% and confidence =65%.</p> <p>The proposed visualization aims to offer a clear representation of rules but also interactive where rules can be explored by zooming and filtering techniques (cf. Fig. 9).</p>   |
| <p>Knowledge management</p> <p>Interacting with the visual data mining tool to visually acquire new knowledge on the patient state with low cognitive load</p> | <p>The physician (user/decision-maker) perceives the visual temporal association rules. He/she interacts within the representation using the specifications described above.</p> <p>Based on his/her experience in nosocomial infection surveillance and the performed training on the visualization tool, the physician can understand and visually interpret the perceived visual elements. Association rules are easy to understand but can become large when automatically generated. The physician can gain a first understanding of the rules by following their associations (understanding by visualizing). A complementary understanding can be performed by exploiting the brain’s ability [24] to: (1) cognitively reason about the rules cause and effect and, (2) let the physician interact with the representation using the specification for understanding by doing.</p> <p>Once comprehended, the physician can trust the visual predicted model based on the support and confidence factors visualization. The knowledge is built and drives interactive exploration for supporting decision on the patient state.</p> <p>The physician enters in the cognitive loop of perception, comprehension, trust, knowledge and exploration. Such loop allows him/her to synthesize the critical relationships and make the best decision at day <math>i</math> (<math>0 \leq i \leq</math> hospitalization stay) to avoid the nosocomial infection occurrence. It is an interlinking dynamic decision-maker-visualization interaction.</p> |

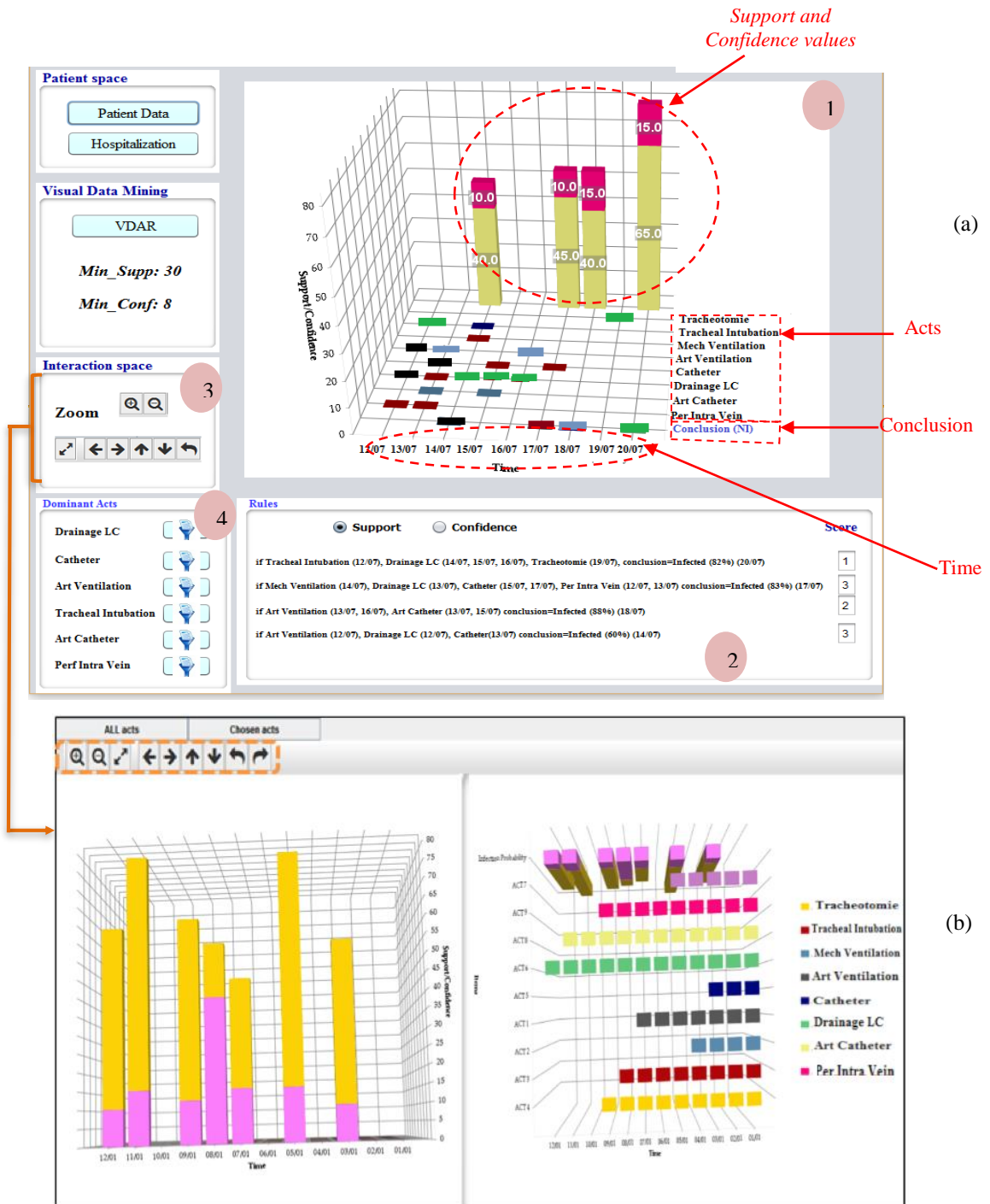


Fig. 9. (a) Visual temporal association rules and (b) applied interaction mechanisms



To visualize dynamic association rules, we applied the enhanced process to develop 3D Histogram as improvement of rules-item matrix (Fig. 9(a)). It increases the intelligibility of dynamic rules by a more detailed representation. It can be very beneficial by providing physicians with meaningful visual elements (Matrix, cube, rectangle...) instead of textual representation, which is poor (e.g. rule browsers [22]). Using 3D Histogram Technique in conjunction with dynamic association rules algorithm accelerates and facilitates data analysis. This technique representing generated rules consists of matrix visualizing acts, conclusions and days of hospitalization, and histograms representing metrics (Support and confidence). In fact, taking an act at a time  $t$  is expressed by the intersection of the X-axis (time) and the Y-axis (acts) (2D matrix). This graphical encoding visualizes each rule (acts and conclusion) using rectangle with specific colour and corresponding metrics are represented by the histogram visualized at the same date. Via this user interface (Fig. 9(b) Dominant act section) the physician can select the dominant act (frequency of the act) to filter items. Eventually, filtering acts allows user to participate in the association rule mining and construct a new representation. Consequently, the number of displayed rules will be reduced then the cognitive load of analysis will be lower. In addition, via rules section, the user can sort the generated rules using support or confidence factor. Then, if the user considers it interesting, he/she attributes it a score (between 1 and 3). These rules, having the highest score, constitute history, which will be used to compare and extract knowledge.

In order to further reduce the cognitive load, interaction tools (Fig. 9(a) Interaction space) like zooming facilitate the understanding of temporal rules. The proposed 3D Histogram Technique provides the ability to interact with temporal association rules algorithm and to interpret and evaluate the rules quality.

The example presented by Fig. 9(a) shows how we can assist decision-making using the developed tool. To do so, Fig. 9(a) (visualization space 1) shows the situation of a patient who takes medical acts daily during hospitalization. 3D Histogram technique representing generated rules, between 12/07/2014 and 20/07/2014, consists of 2D matrix visualizing acts, conclusions and days of hospitalization, and histograms representing metrics (Support and confidence). This technique describes data relationships in the form (If-Then) rules among different acts and NI probabilities during hospitalization of selected patient (using Patient Data button). As an example of generated rules visualized using 3D Histogram we can cite "if applied acts are: tracheal intubation in 12/07, drainage LC in 14, 15, 16 and tracheotomie in 19/07 then patient in 20/07 is 0.82 infected". To visualize it, we use green rectangle in the intersection of acts and result axe and time axe, result value (infected) are also represented using the same colour. In the same space and using histograms, we represent measures of association rules evaluation, which are the Support and the Confidence by indicating the values (15% and 65% for the cited rule). The important factors here are support and confidence. We propose to filter rules using minimum threshold ( $\text{minsupp}=30$  and  $\text{minconf}=8$ ). Via this user interface, Fig. 9(a) (visualization space 2), we represent temporal rules in textual representation that can be sorted by confidence or support values. In this space, we give the hand to decision maker to attribute a score between 1 and 3 for each rule. Then, rules having the highest score will be stored to make the patient's history. Those stored rules are useful for knowledge generation because it represents the most pertinent for the user. For this reason, he/she can gain insight it for generating new knowledge. For further interaction human-visualization technique, decision-maker can use the interaction space (visualization space 3) consisting of zoom tool and rotation tools (left, right, up or down). As shown in Fig. 9(b), after applying right and up rotations, user can see more details for example to facilitate the correspondence between acts, the results and the hospitalization period. In addition, to display rules containing a specific act, decision-maker can use visualization space 4 and select the desired act. After representing related rules, physician can determinate very fast the effect of using the selected act. For this reason, each interaction tool is useful in generating and enhancing knowledge by exerting a low cognitive load. For example, if we choose Catheter then just the rules having this act as antecedent will be represented. Consequently the decision-maker can determine if it is the principal reason of

nosocomial infection appearance. Finally, we note that the tool integrated in our 3D representation helps to generate new knowledge, as a direct result of assisting the physician to make the suitable decision for the patient's situation.

As explained before, we used a minimum threshold filter rules in order to visualize pertinent information. However, the algorithm of prediction can give unexpected, contradictory or unsatisfying result, then user attribute as score of interest the lowest value using space of visualization 2. Each rule can be validated by assigning high, medium or lower score. This score will be stored, if the score is equal to 1 then the combination of acts used will be eliminated without calculating support, confidence and probability of catching NI.

Finally, we note that explaining the patient's situation (infected/not infected) dynamically using temporal visual representation makes the thought process faster and more conscious. Thus, the task of decision making benefits from visual explanation (3D representation) and from interaction tools (filter, rotation, zoom, labeling). Consequently, the cognitive load exerted for understanding, analyzing visual association rules in order to get new knowledge useful for dynamic decision-making is decreased significantly.

## **5. Evaluation**

The visual temporal association rules tool has been evaluated by the ICU physicians to assess its utility and usability [39] [52]. The participants' profiles are as follows:

1. Nosocomial Infection domain expert: two physicians with an average age of 40 years.
2. Nosocomial Infection domain novice: two Health Informatics (HI) Professionals with an average age of 35 years,
3. Software developer: two PhD students of 26 years old, volunteered from the university community who are not familiar with the nosocomial infection concepts but they are expert in information technology and data mining.

All user profiles are familiar and easily use computers. In this section, we present the evaluation protocol and results.

### *5.1. Evaluation of the prediction results*

The developed temporal association rules algorithm was applied on a test base of 80 ICU patients. To verify its performance, we have used the confusion matrix (Table 2). Each column of the matrix lists the instances in a predicted class while each row represents the instances in an actual class.

This table allows detailed analysis of the prediction ability of the algorithm. The associated performance metrics [54] were calculated; they are presented in the table 2.

Table 2. Confusion matrix

|               |          | Predicted values        |                          | Total |
|---------------|----------|-------------------------|--------------------------|-------|
|               |          | Positive                | Negative                 |       |
| Actual values | Positive | True Positives (TP)= 18 | False Positives (FP) =11 | 29    |
|               | Negative | False Negatives (FN) =8 | True Negatives (TN) =43  | 51    |
|               | Total    | 26                      | 54                       | 80    |

**Performance metrics**

$$\text{Sensitivity or True Positive Rate (TPR)} = TPR = \frac{TP}{P} = \frac{TP}{TP + FN} = 69\%$$

$$\text{Specificity (SPC) or True Negative Rate (TNR)} = SPC = \frac{TN}{N} = \frac{TN}{FP + TN} = 79\%$$

$$\text{Precision or Positive Predictive Value (PPV)} = PPV = \frac{TP}{TP + FP} = 62\%$$

$$\text{Negative Predictive Value (NPV)} = NPV = \frac{TN}{TN + FN} = 84\%$$

$$\text{Fall-out or False Positive Rate (FPR)} = FPR = \frac{FP}{N} = \frac{FP}{FP + TN} = 1 - SPC = 21\%$$

$$\text{False Discovery Rate (FDR)} = FDR = \frac{FP}{FP + TP} = 1 - PPV = 38\%$$

$$\text{Miss rate or False Negative Rate (FNR)} = FNR = \frac{FN}{P} = \frac{FN}{FN + TP} = 30\%$$

$$\text{Accuracy (ACC)} = ACC = \frac{TP + TN}{P + N} = 76\%$$

F1 score: it is the harmonic mean of precision and sensitivity =

$$F1 = \frac{2TP}{2TP + FP + FN} = 65\%$$

The calculated performance rates visible in the table 2 show that the developed tool provides encouraging experimental results. In fact, the predictive tool is accurate at 76 %. The precision rate of 62% means that the dynamic algorithm returned significantly more relevant results than irrelevant ones, while the sensitivity rate of 69% means that it returned most of the relevant results. The weighted average of the precision and sensitivity is equal to 65% (F1 score), which is considered as interesting predictive value and suggests a preliminary validation of the tool.

Table 3 shows that the utility performance is better than the results obtained in the two previous versions of our system presented in [6] and [42]. In the first version [6], we have proposed a user-centered approach, called UP/U, based on the Unified Process from the field of the Software Engineering and the U model from the field of the Human-Computer Interaction to develop KDD-based DSS in static environment. We have developed the case-based reasoning technique (K Nearest Neighbours (KNN)). This first system was accepted by the end users (cf. table 3). Concerning the second version [42], we have proposed the Extended Unified Process as a new method to design and develop a KDD-based DSS in dynamic environment, by involving the

end user all through the development process in order to make it possible for potential users to describe their functional needs and to evaluate and validate the different interfaces. The developed data mining technique was the Dynamic Bayesian Networks taking into account the temporal characteristic of data. Utility evaluation rates are measured and have shown good results (cf. table 3).

Table 3: comparison of the different versions of the intelligent dynamic DSS

|   | Contribution   | Utility evaluation   |
|---|--|--|
| <b>The first version of DSS [6]</b>           | Data mining technique: case-based reasoning  | The decision-makers (ICU physicians) were involved in the human-centered development process. The evaluation results are not measured but expressed only in terms of system acceptance and validation. |
| <b>The second version of dynamic DSS [42]</b> | Data mining technique: Dynamic Bayesian Networks   | <ul style="list-style-type: none"> <li>- The accuracy rate = 74%</li> <li>- The error rate = 26%</li> <li>- The sensibility rate = 56%</li> <li>- The specificity rate= 81%</li> </ul>                 |
| <b>The Current version of dynamic DSS</b>     | <ul style="list-style-type: none"> <li>- Data mining technique: Dynamic association rules</li> <li>- Temporal data manipulation</li> <li>- Temporal visualization</li> <li>- Cognitive knowledge management</li> </ul> | <ul style="list-style-type: none"> <li>- The accuracy rate = 76%</li> <li>- The error rate = 21%</li> <li>- The sensibility rate = 69%</li> <li>- The specificity rate = 79%</li> </ul>                |

Based on the comparison presented in table 3, we can deduce that the prediction rates are improved using the developed decisional tool. After that, we will also investigate its usability.

## 5.2. Usability evaluation

To evaluate the usability of the visualization tool, visible in Fig. 9, we have used the insight based evaluation method of [57]. An insight is defined as an individual observation about the data by the participant. Since the primary purpose of the visual tools is to provide relevant insight into the data, this method aims to evaluate visual representations based on numerous characteristics of insight that enabled to recognize and quantify it in open-ended user tests. We select this method (1) to understand how our visualization tool can support decision-makers analysis and reasoning; and (2) to understand how they can communicate about their insights using such visualization.

The main insight characteristics as introduced by [57] are: *time*, *hypotheses*, *directed versus unexpected*, *correctness* and *category* (overview (overall distributions of data), patterns (identification or comparison across data attributes), *groups* (identification or comparison of groups of data), and *details* (focused information about specific data)).

### 5.2.1. Evaluation protocol

Six participants (whose profiles have been provided above) were volunteers to assess the usability of the tool. A training session was planned to: (1) describe the temporal data set and, (2) to demonstrate the tool usage and interactivity during a 20-minute tutorial. We have measured the time spent by each user to learn the tool during the training session.

To evaluate the tool in terms of its ability to generate insight, the protocol consists of identifying the individual insight occurrences while participants analyse the temporal data (cf. Fig. 10).

Participants listed a set of analysis questions they would ask about antibiotics, acts, biological parameters and infectious examinations data sets. Then, they continue to test the data with the tool until they decided that

they would not gain any further insight. The training session was videotaped to later identify the individual occurrences of insights.

During participants' work, they were asked to:

1. Comment their observations and conclusions.
2. Estimate, every 15 minutes, the total of potential insight they had learned about the temporal data.
3. Present any difficulties or benefits of the tool at the end of the session period.

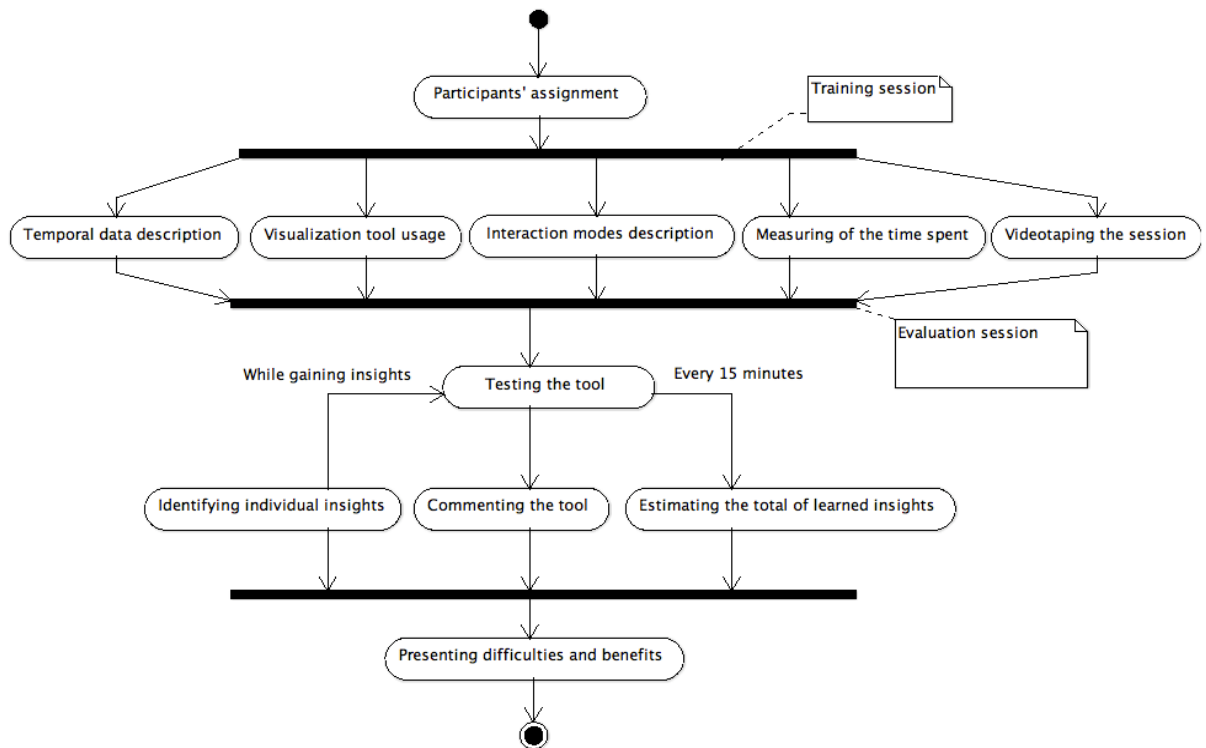


Fig. 10. The evaluation protocol

### 5.2.2. Insight-based evaluation

To evaluate the visual data mining tool for the decision-making basing on the insight characteristics listed above, we have first counted the total number of insights (distinct observations about the temporal ICU data to analyse) by each participant. The group who analysed the same temporal data set reported very similar insights about the data. Consequently, the reported insights were repetitive across participants. We found that the total count was 25.

The table 4 summarizes the total number of generated hypotheses, unexpected insights, and incorrect insights per participant.

Table 4. Insight characteristics

| Participant       | Insights characteristics |                     |                    |
|-------------------|--------------------------|---------------------|--------------------|
|                   | Hypotheses               | Unexpected insights | Incorrect insights |
| Physician 1       | 2                        | 3                   | 2                  |
| Physician 2       | 3                        | 5                   | 1                  |
| HI professional 1 | 2                        | 3                   | 1                  |
| HI professional 2 | 1                        | 4                   | 1                  |
| PhD student 1     | 0                        | 0                   | 1                  |
| PhD student 2     | 0                        | 1                   | 1                  |

From the table 4, we generated the histogram visible in the Fig. 11.

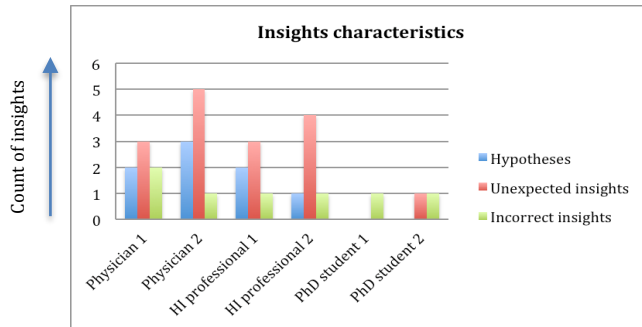


Fig. 11. The insights characteristics evaluation

According to the results presented in table 4 and Fig. 11, only a few insights led participants to new medical hypotheses. For instance, the second physician commented that parts of the time series data showed a regular behaviour. He searched for antibiotics, acts and germs association that showed similar behaviour at earlier time points, but could not find any. He offered several possible explanations for this behaviour in relation with the NI occurrence.

The temporal association rules tool provided two unexpected insights from the ICU data set and one unexpected insight about the time series data. This visual tool proved helpful to users. However, the first physician reported an incorrect insight due to inferring the colour scale of the 3D histogram. The second incorrect insight was reported by all the participants. It is related to the cluttered representation of a large set of generated rules at a specific point of time. Future development will focus on enhancing the visual temporal association rules tool with advanced interactive features in addition to filtering, zooming and labelling.

Table 5 presents the characteristics of the insight categories and the time. Concerning the time, we have counted: (1) the time for first insight occurrence of each participant where lower times indicate that participants are able to get immersed in the temporal data more quickly and (2) the total time that each participant spent using the tool until he/she felt he/she could gain no more insight. These two characteristics summarize the time to acquire insights.

Table 5. Insights per participant

| Participant       | Insight categories |          |        |         | Time                   |            |
|-------------------|--------------------|----------|--------|---------|------------------------|------------|
|                   | Overview           | Patterns | Groups | Details | Time for first insight | Total time |
| Physician 1       | 13                 | 10       | 2      | 5       | 4 min                  | 26 min     |
| Physician 2       | 10                 | 8        | 4      | 3       | 6 min                  | 28 min     |
| HI professional 1 | 6                  | 7        | 0      | 2       | 5 min                  | 35 min     |
| HI professional 2 | 9                  | 10       | 0      | 4       | 6 min                  | 32 min     |
| PhD student 1     | 5                  | 8        | 2      | 1       | 2 min                  | 19 min     |
| PhD student 2     | 7                  | 6        | 1      | 2       | 3 min                  | 20 min     |

From the table 5, we generated the histogram visible in the Fig. 12.

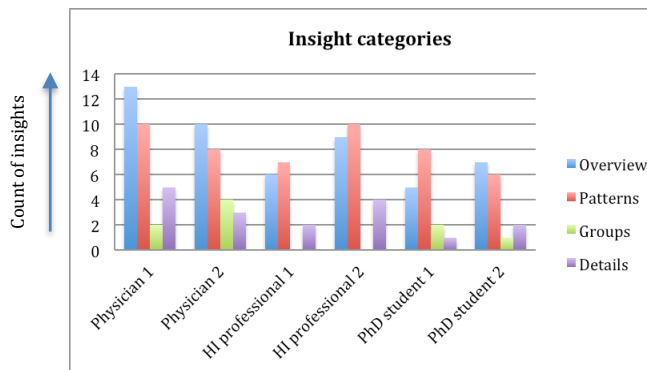


Fig. 12. The insights categories evaluation

Interpreting table 5 and Fig. 12 reveals that participants (P1 to P6) with more domain expertise compared to others would gain more insight from the temporal data using the dynamic visual association rules mining tool, which is very encouraging. In addition, we found that the number of insights were dependent on the participant profile but also on their motivation. Considering the factor of time for first insight, we found that the software developers (participants P5 and P6) made the insights faster than other participants due to their computer experience.

Based on this insight evaluation, a set of other improvement suggestions of our prototype was proposed by user profile:

1. The physicians (participants P1 and P2) underlined their desire to directly select the temporal attributes in the tool by a pull-down menu or by specific 3D graphics objects.
2. The HI Professionals (P3 and P4) suggested improving the efficiency of our system by other means of interactive knowledge extraction techniques that have not been implemented due to lack of time and still require long time of development.
3. The PhD students (P5 and P6) proposed to graphically encode additional interest measures (e.g. calculate the informational gain and materialize the involvement of intensity between the premise and the conclusion of the visualized rule).

Concerning the interaction modes of our tool, we found that most of the physicians and HI professionals preferred the filtering function of the 3D visual representation, whereas PhD students preferred the zoom view. This is despite the fact that the two first user profiles are domain experts and find filtering helpful in

visually generating knowledge. Most of these participants performed the same analyses using different views. They showed, using the interaction modes, a higher performance on displaying and interpreting data sets. However, they believed that it would be more useful to them if the interaction capabilities of this tool were increased, e.g., to better subdividing data, enabling comparison, etc.

Finally, we note that the evaluation conclusions have proven that our system allowed users to visually discover knowledge from temporal data easily and with limited cognitive load. Hence, the proposed improvements show that there is some work on the interactive and encoding aspects of the tool.

## **6. Discussion**

This paper presented the design of visual data mining method for dynamic decision-making. Such design must follow a visualization process taking into account: data mining, temporal and cognitive considerations. In this context, we have proposed to adapt the visualization process of [72] under these considerations (cf. §3, Fig. 6). Our proposal structures the adapted visual data mining process into 3 steps: (1) temporal data manipulation, (2) temporal visualization, and (3) discovered knowledge management.

The basic contribution of this work is the proposed adapted visual data mining process to design and develop visual intelligent dynamic DSS representations. On the methodological side particularly aiming at designing visual intelligent dynamic DSS, this current work consists of integrating temporal data mining specificities in an existing visualization process while decoding the decision-makers behaviours. Hence, it seems essential for accurate visual data mining representations in dynamic decision-making activity to propose a cognitive temporal visual data-mining pipeline that allows comprehending the cognitive aspects of this activity.

This study identifies new perspectives on intelligent dynamic DSS. Our pipeline highlights that human and machine are a loop in the temporal knowledge generation and integration process for decision-making. While existing visual data mining processes focus on one of these, our process integrates human thinking and temporal aspects whilst describing visual data mining based DSS components.

Most of the existing intelligent DSS do not fully cover all the aspects that our process requires. In fact, it includes the whole process of knowledge extraction and integration for dynamic decision-making, which involves visual data mining tasks and human decision-makers in dynamic loop. Our process specifies the whole path from goals definition to decision-making including visual and temporal analytics tasks and vice versa. Besides connecting to system components, the process also defines human concepts and introduces cognitive reasoning.

As a practical implication, the development of the visual temporal association rules based on our enhanced process was concretely applied for the fight against ICU nosocomial infections. The progress of the proposed process is presented by table 1 showing the contribution of the temporal cognitive data mining specificities. The significant practical implication of this work is the final product under use in the ICU of Habib Bourguiba Hospital of Sfax, Tunisia. In fact, this latter is now used to visually prevent nosocomial infections, for each patient, using the temporal association rules mining algorithm (in this paper we have not focused on the algorithmic contribution of our tool). The performed utility and usability evaluation allowed to validate the prediction capacity as well as the usability of the generated association rules representations, except some errors in relation with displaying large sets of rules. Several improvements have been suggested by the users in relation with interactive features and interest measures of association rules.

An interesting work of Zhou and Mabou [75] proposed a method using Genetic Network Programming (GNP). This work aims to generate time related association rules using database consisting of a big amount of time related attributes. Their method works well in dynamic environments especially to solve the road traffic prediction problem. Besides, it requires domain knowledge, there is no user intervention in the data mining



process and they did not use visual representation to improve the computer process; whereas, in the previous section, we demonstrated the importance of using visualization and of integrating user in the data analysis process.

All proposed techniques based on rules's visualization represent an association rule as a whole without taking into consideration the relations among items. Ben Said et al. [8] proposed a visualization technique for association rules that represent antecedent, consequent and links between each pair of them. Using this technique, the user is able to be involved in the algorithm of rule extraction, and not just to manipulate the rule visualization. This is a common point processed in our work.

This discussion and previous evaluation reveal research challenges in advancing our visual temporal analytic method. Especially, we note that there is no incorporation of semantic knowledge about temporal association rule models in the visualization process; this would be a relevant research way.

## **7. Conclusion and further study**

This research aims at designing and developing visual data mining for supporting decision-making in dynamic situations. We intend to attain more effective and human-perceptive dynamic visualizations. To do so, we have proposed the enhancement of an existing visual analytics process [72] by adapting and restructuring it under three points: the temporal dimension, the data mining specificities and the cognitive considerations.

We have illustrated the cognitive temporal visual data mining process by considering the design of the visual and temporal association rules mining. The developed tool has been then evaluated in terms of utility to assess its predictive ability and usability to assess its ability to generate insight. The evaluation results are encouraging and have motivated us: (1) to meet users' proposal for the prototype improvement, and (2) to apply the proposed process on other visual data mining techniques such as the Dynamic Bayesian networks.

What we have introduced in this paper is an enhanced process still in its primary stages, whose future will absolutely be improved by additions, corrections and additional series of experimental evaluations. We plan also to adapt such visual data mining process to mobile environments. In fact, mobile computing and communication provides an exciting opportunity for visual data mining based DSS in dynamic environments. These systems can be very beneficial to specific application domains where complex and critical decisions must be made under time pressure (such as healthcare, emergency management, commerce and banking applications) [7] [11]. Besides, adapted visual data mining on mobile devices can significantly support mobility of decision-makers by accessing and/or analysing data where and when it is needed the most

## **Acknowledgements**

The authors would like to acknowledge the financial support of this research by grants from the ARUB program under the jurisdiction of the General Direction of Scientific Research (DGRST) (Tunisia). Thanks are also due to all the ICU staff of Habib Bourguiba Teaching Hospital for their interest in the project and all the time they spent helping us design, use and evaluate our system.

## **References**

- [1] Aigner, W., Miksch, S., Muller W. et al., 2007, Visualizing Time-Oriented Data – A Systematic View, *Computers & Graphics*, vol. 31(3), pp. 401- 409.
- [2] Aigner, W., Federico, P., Gschwandtner, T., Miksch, S. and Rind A., 2012, Challenges of Time-oriented Data in Visual Analytics for Healthcare, *IEEE VisWeek Workshop on Visual Analytics in Healthcare*, IEEE.

- [3] Augusto, J.C. 2005, Temporal reasoning for decision support in medicine, *Artificial Intelligence in Medicine*, vol. 33(1), pp. 1–24.
- [4] Basole, R.C., Kumar, V., Park, H., Braunstein, M.L., Kahng, M., Chau, D.H., Hirsh, D.A., Serban, N., Bost, J., Lesnick, B., Schissel, B., Tamersoy, A., and Thompson M., 2015, Understanding variations in pediatric asthma care processes in the emergency department using visual analytics, *Journal of the American Medical Informatics Association*, vol. 22 (2), pp. 318-323.
- [5] Batagelj, V., Brandenburg, F.J., Didimo, W., et al., 2011, Visual analysis of large graphs using (X,Y)-clustering and hybrid visualizations. *IEEE Transactions on Visualization and Computer Graphics*, vol. 17 (11), pp. 1587–1598.
- [6] Ben Ayed, M., Ltfi, H., Kolski, C., Alimi, M-A., 2010, A User-centered Approach for the Design and Implementation of KDD-based DSS: A case Study in the Healthcare Domain. *Decision Support Systems*, 50 (1), pp. 64-78.
- [7] Ben Ayed, E., Ben Ayed, M., Kolski, C., Ezzedine, H., Gargouri, F., 2015, Context Aware Criteria for the Evaluation of Mobile Decision Support Systems. *ICIS 2015, 14th IEEE/ACIS International Conference on Computer and Information Science (June 28 - July 1)*, IEEE, Las Vegas, USA, pp. 661-666.
- [8] Ben Said, Z., Guillet, F., Richard, P., and Picarougne F.M., 2013, Visualization of association rules on a molecular representation. In: *The 17th international conference on Information Visualization*.
- [9] Ben Mohamed, E., Ltfi, H., and Ben Ayed, M., 2013, Using visualization techniques in knowledge discovery process for decision-making. In: *The 13th international conference on hybrid intelligent systems (HIS)*, Tunisia, pp. 94–99, New York: IEEE.
- [10] Burnett, C., 2011, *Trust Assessment and Decision-Making in Dynamic Multi-Agent Systems*, Ph.D. Dissertation, Department of Computing Science, University of Aberdeen,
- [11] Burstein, F., and Cowie, J., 2008, Mobile decision support for time-critical decision-making, In *Encyclopedia of Decision Making and Decision Support Technologies*, Edited by: Adam, Frédéric and Humphreys, pp. 638–644, Hershey, PA: Information Science Publishing, Hershey.
- [12] Candelieri, A., Dolce, G., Riganello, F., and Sannita, W.G., 2011, Data Mining in Neurology, In *Knowledge-Oriented Applications in Data Mining*, Kimito Funatsu (Ed.), pp. 261-276, InTech.
- [13] Cannon-Bowers, J.A., Salas E. and Converse, S.A., 1993, Shared mental model in expert decision making teams, In N. J. Castellan Jr. (Ed.), *Current Issues in Individual and Group Decision Making*, pp. 221-246, Hillsdale, NJ: Lawrence Erlbaum Associates.
- [14] Chi Ed. H., 2000, A Taxonomy of Visualization Techniques using the Data State Reference Model, In *INFOVIS '00 proceedings of the IEEE Symposium on Information Visualization 2000*, pp. 69 – 75, IEEE Computer Society Washington, DC, USA.
- [15] Dearden, A., Harrison M., and Wright, P., 2000, Allocation of function: scenarios, context and the economics of effort, *International Journal of Human-Machine Studies*, vol. 52, pp. 289-318.
- [16] Daassi C., Nigay L., Fauvet M-C., (2004). Visualization process of Temporal Data, In *proceedings of DEXA conference*, 914-924, Zaragoza, Spain, LNCS 3180, Springer.
- [17] Dereli, N., Ozayar E., Degerli S., Sahin S., Koç F., 2013, Three-year evaluation of nosocomial infection rates of the ICU, *Braz J Anesthesiol.*, vol. 63(1), pp. 73-78.
- [18] El-Masri, M.M., and Oldfield, M., 2012, Exploring the influence of enforcing infection control directives on the risk of developing healthcare associated infections in the intensive care unit: A retrospective study, *Intensive and Critical Care Nursing*, vol. 28(1), pp. 26-31.
- [19] Fayyad, U.M., Piatetsky-Shapiro, G., and Smyth, P., 1996, From Data Mining to Knowledge Discovery in Databases, *AI Magazine*, vol. 17(3), pp. 37-54.
- [20] Fayyad, U.M., Wierse, A. and Grinstein, G., 2002, *Information visualization in data mining and knowledge discovery*, Morgan Kaufmann Publishers Inc., San Francisco, CA.
- [21] Fki, H., Yaich, S., Jdidi, J., Karray, A., Kassis, M., Damak, J., 2008, Epidemiologie des infections nosocomiales dans les hopitaux universitaires de sfax: résultats de la première enquête nationale de prevalence de l'infection nosocomiale, *Rev. Tun. Infectiol.*, vol. 2(1), pp. 22-31.
- [22] Fule, P., and Roddick, J., 2004, Experiences in building a tool for navigating association rule result sets. In *Proceedings of ACSW Frontiers*, volume 32, page 103-108.
- [23] Funatsu, K., 2011, *New Fundamental Technologies in Data Mining*, 596 pages, Publisher: InTech.
- [24] Glimcher, P-W., and Fehr, E., 2013, *Neuroeconomics: Decision making and the brain*, 2nd edition, Academic Press, 560 pages.
- [25] Han, J., Kamber, M., and Pei, J., 2006, *Data Mining: Concepts and Techniques*, Morgan Kaufmann; 2nd edition, 800 pages.
- [26] Hibino, S., Rundensteiner, E.A., 1998, Comparing MMVIS to a timeline for temporal trend analysis of video data, In *AVI '98 Proceedings of the working conference on Advanced visual interface*, pp. 195 – 204, ACM New York, NY, USA.
- [27] Hoc, J.M., and Amalberti, R., 1995, Diagnosis: some theoretical questions raised by applied research, *Current Psychology of Cognition*, vol. 14, pp. 73-100.
- [28] Holzinger, A., 2013, Human–computer interaction & knowledge discovery (HCI-KDD): what is the benefit of bringing those two fields to work together? In: C.K. Alfredo Cuzzocrea, D.E. Simos, E. Weippl, et al. (eds.), *Multidisciplinary research and practice for information systems (Springer Lecture Notes in Computer Science LNCS 8127)*. Heidelberg, Berlin and New York: Springer, pp. 319–328.

- [29] Holzinger, A., Dehmer, M., and Jurisica, I., 2014, Knowledge Discovery and interactive Data Mining in Bioinformatics - State-of-the-Art, future challenges and research directions, BMC Bioinformatics, vol. 15, iss. S6, p. 11.
- [30] Jungerman, H., and Fischer, K., 2005, Using expertise and experience for giving and taking advice, in T. Betsch, and S. Haberstroh (eds.), The Routines and Decision Making, Lawrence Erlbaum, Mahwah, pp. 157- 173.
- [31] Kammoun, F. and Ben Ayed, M., 2014, Clinical Dynamic Decision Support System Based on Temporal Association Rules, Middle East Conference on Biomedical Engineering (MECBME) February 17-20, Doha, Qatar, pp. 289-292.
- [32] Keim, D., 2002, Information Visualization and Visual Data Mining, IEEE Transactions on Visualization and Computer Graphics, vol. 7, no 1, pp. 100-107.
- [33] Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J., and Melancon, G., 2008, Visual Analytics: Definition, Process, and Challenges. In Information Visualization, A. Kerren, J.T. Stasko, J-D. Fekete, and C. North (Eds.). Lecture Notes In Computer Science, Vol. 4950. Springer-Verlag, Berlin, Heidelberg, pp. 154-175.
- [34] Kowalczyk, Z., and Czubenko, M., 2010, Interactive Cognitive-Behavioral Decision Making System, Artificial Intelligence and Soft Computing chapter, In 10th International Conference, ICAISC 2010, Zakopane, Poland, June 13-17, 2010, Part II, Volume 6114 of the series Lecture Notes in Computer Science, pp. 516-523.
- [35] Kumar, V., Park, H., Basole, R.C., 2014, Braunstein M, Kahng M, Chau DH, Tamersoy A, Hirsh DA, Serban N, Bost J et al (2014) Exploring clinical care processes using visual and data analytics: challenges and opportunities, In Proceedings of the 20th ACM SIGKDD conference on knowledge discovery and data mining, workshop on data science for social good.
- [36] Le Coze, J.C., 2015, Reflecting on Jens Rasmussen's legacy, A strong program for a hard problem, Safety Science, Vol. 71, Part B, pp. 123-141.
- [37] Lee, J.D., Moray, N., 1994, Trust, self-confidence, and operators' adaptation to automation, International journal of human-computer studies, vol. 40 (1), pp. 153-184.
- [38] Lewis, J.D., and Weigert, A., 1985, Trust as a social reality. Social Forces, vol. 63(4), pp. 967-985.
- [39] Lewis, J.R., 2014, Usability: Lessons Learned and Yet to Be Learned. International Journal of Human-Computer Interaction , 30 (9), pp. 663-683.
- [40] Li, H., and Yang, L. 2013, Time series visualization based on shape features, Knowledge-Based Systems, Vol. 41, pp. 43-53
- [41] Linoff, G.S., Berry, M.J.A., 2011, Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management, 3rd Edition, Wiley, 888 pages.
- [42] Ltifi, H., Kolski, C., Ben Ayed, M., and Alimi, A.M., 2013, Human-centered Design Approach for Developing Dynamic Decision Support System based on Knowledge Discovery in Databases, Journal of Decision Systems, vol. 22(2), pp. 69-96.
- [43] Ltifi, H., Trabelsi, G., Ben Ayed, M., and Alimi, A.M., 2012, Dynamic Decision Support System Based on Bayesian Networks, application to fight against the Nosocomial Infections, International Journal of Advanced Research in Artificial Intelligence (IJARAI), Vol. 1, no. 1, pp. 22-29.
- [44] Ltifi, H., Kolski, C., Ben Ayed, M., 2015, Combination of cognitive and HCI modeling for the design of KDD-based DSS used in dynamic situations. Decision Support Systems, 78, pp. 51-64.
- [45] Ltifi, H., Ben Mohamed, E., and Ben Ayed, M., Interactive visual KDD based temporal Decision Support System, Information Visualization, Vol. 15, no. 1, pp. 31-50, 2016.
- [46] Mackinlay J.-D., Robertson, G.G., and Card, S.K., 1991, The Perspective Wall: Detail and Context Smoothly Integrated. Proceedings of the CHI'91, ACM Press, pp. 173-179.
- [47] McNaught, K.R., and Zagorecki, A., 2009, Using dynamic Bayesian networks for prognostic modelling to inform maintenance decision making, IEEE International Conference on Industrial Engineering and Engineering Management, IEEM 2009, pp. 1155 – 1159, Hong Kong, IEEE.
- [48] Millot, P., 2014, Cooperative organization for Enhancing Situation Awareness, In P. Millot (Ed.), Risk Management in Life critical Systems, ISTE-Wiley, London, pp. 279-300.
- [49] Milovic, B., and Milovic, M., 2012, Prediction and decision-making in Health care using data mining, Kuwait Chapter of Arabian Journal of Business and Management Review, Vol. 1, No.12, pp. 126-136.
- [50] Miner, L., Bolding, P., Hilbe, J., Goldstein, M., Hill, T., Nisbet, R., Walton, N., and Miner, G., 2014, Practical Predictive Analytics and Decisioning Systems for Medicine, Elsevier Inc., p. 1110.
- [51] Mookiah, M.R.K., Acharya, U.R., Lim, C.M., Petznick, A., and Suri J.S., 2012, Data mining technique for automated diagnosis of glaucoma using higher order spectra and wavelet energy features, Knowledge-Based Systems, Volume 33, pp. 73-82.
- [52] Nielsen, J., 1993, Usability engineering, Morgan Kaufmann, ISBN 0-12-518406-9, San Francisco, USA, 362 pages.
- [53] Perl, T.M., 1997, Surveillance, reporting and the use of computers. In: Prevention and control of nosocomial infections, third edition. RP Wenzel, ed. Baltimore, Williams & Wilkins, 1997, pp. 127-161.
- [54] Powers, D., 2011, Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation, Journal of Machine Learning Technologies, vol. 2 (1), pp. 37-63.

- [55] Rasmussen, J., 1986, *Information processing and human-machine interaction, an approach to cognitive engineering*, Elsevier, Amsterdam.
- [56] Rind, A, Wang, T-D, Aigner, W, Miksch, S., Wongsuphasawat, K., Plaisant, C., and Shneiderman, B., 2013, Interactive information visualization to explore and query electronic health records, *Found Trends Hum-Comput Interact*, vol. 5(3), pp. 207–298.
- [57] Saraiya, P., North, C., and Duca, K., 2005, An Insight-Based Methodology for Evaluating Bioinformatics Visualizations, *IEEE transactions on visualization and computer graphics*, vol. 11 (4), pp 443-456.
- [58] Salas, E., Prince, C., Baker, D.P., and Shrestha, L., 1995, Situation awareness in team performance: implications for measurement and training, *Human Factors*, vol. 37, pp. 123-136.
- [59] Shen, B., Yao, M., Wu, Z., and Gao, Y., 2010, Mining dynamic association rules with comments, *Knowledge and Information Systems*, Vol. 23 (1), pp. 73-98.
- [60] Simoff, S.J., Böhlen, M.H., and Mazeika, A., 2008, *Visual Data Mining, Theory, Techniques and Tools for Visual Analytics*, Lecture Notes in Computer Science 4404, Springer Berlin Heidelberg.
- [61] Smyth, E.T., McIlvenny, G., Enstone, J.E., Emmerson, A.M., Humphreys, H., Fitzpatrick, F., Davies, E., Newcombe, R.G., and Spencer, R.C., 2008, Four Country Healthcare Associated Infection Prevalence Survey 2006: overview of the results, *Journal of Hospital Infection*, vol. 69(3), pp. 230-248.
- [62] Suetens, C., Morales, I., Savey, A., Palomar, M., Hiesmayr, M., Lepape, A., Gastmeier, P., Schmit, J.C., Valinteliene, R., Fabry, J., 2007, European surveillance of ICU-acquired infections (HELICS-ICU): methods and main results, *J Hosp Infect.*, 65 Suppl 2, pp. 171-173.
- [63] Thearling K., Becker B., DeCoste D., Mawby B., Pilote M., and Sommerfield D., *Visualizing Data Mining Models*, 2001, *Information Visualization in Data Mining and Knowledge Discovery*, Eds Fayyad U., Grinstein G., and Wierse A., Morgan Kaufman.
- [64] Thomas, J., Cook, K. (Eds), 2005, *Illuminating the Path: Research and Development Agenda for Visual Analytics*, Los Alamitos, CA, IEEE CS-Press.
- [65] Thomas, J.J., and Cook, K.A., 2006, A Visual Analytics Agenda, *IEEE Computer Graphics & Applications*, vol. 26, pp. 10-13.
- [66] Thomas J.J. 2007, Visual analytics: a grand challenge in science: turning information overload into the opportunity of the decade, *Journal of Computing Sciences in Colleges*, vol. 23, pp. 5-6.
- [67] Tran, T.T., 2010, Protecting buying agents in e-marketplaces by direct experience trust modelling, *Knowledge and Information Systems*, vol. 22(1), pp. 65–100.
- [68] Valenzuela, O., Rojas, F., Ortuño, F., Bernier, J-L., Saez, M-J., San-Roman, B., Herrera, L.J., Guillen, A., and Rojas, I., 2014, Advanced Soft-Computing techniques and Clustering Algorithm for Gene Expression Microarray Data Classification, In proceedings IWBBIO 2014, Granada 7-9 April, pp. 1634- 1643.
- [69] Velu, C.M., and Kashwan, K.R., 2012, Performance Analysis for Visual Data Mining Classification Techniques of Decision Tree, Ensemble and SOM, *International Journal of Computer Applications*, vol. 57 (22), pp. 65-71.
- [70] Wang, Y., Ruhe, G., 2007, The Cognitive Process of Decision Making, *International Journal of Cognitive Informatics and Natural Intelligence*, Vol. 1(2), pp. 73-85.
- [71] Wibisono, A., Jatmiko, W., Wisesa, H.A., Hardjono, B., and Mursant, P., 2016, Traffic big data prediction and visualization using Fast Incremental Model Trees-Drift Detection (FIMT-DD), *Knowledge-Based Systems*, Vol. 93(1), pp. 33-46.
- [72] Wijk, V.J.J., 2005, The value of visualization, *Proceedings of IEEE Visualization*, pp. 79 – 86.
- [73] Wong, P.C., and Thomas, J.J., 2004, Visual Analytics, *IEEE Computer Graphics & Applications*, vol. 24, pp. 20-21.
- [74] Yu H., Shen, Z., Leung C., Miao C. and Lesser V., 2013, A survey of multi-agent trust management systems, *IEEE Access*, vol. 1(1), pp. 35–50.
- [75] Zhou, H., Mabu, S., 2009, Generalized time related sequential association rule mining and traffic prediction, *IEEE Congress on Evolutionary Computation (CEC 2009)*.