Abstract—In clinical practice, patient care flows are generally subject to recommended and standardized therapeutic interventions. Especially in a home care setting, situation-aware adherence to therapy can be both difficult for the patient to follow and difficult for the physician to assess. Process mining techniques may be useful artificial intelligence solutions for remotely assessing the compliance of patients' behavior with the corresponding care path, especially if adopted in a cognitive IoT Edge infrastructure, dedicated to the acquisition and analysis of daily routines in a form of event log. In this paper, we present an innovative method to measure in-home adherence to metabolic syndrome management with the aim of providing awareness of the patient's current situation. The analytical results demonstrate the validity of using process mining techniques to remotely evaluate patient behavior.

Index Terms—Situation-aware IoT, Cognitive IoT Edge Compute Systems, Artificial Intelligence, Process Mining, eHealth

I. INTRODUCTION

New frontiers of healthcare delve deeper into eHealth practices, leading to any diagnostic, therapeutic, or social support service provided at home. In fact, home care, when used appropriately, reduces hospitalization and nursing home use without compromising medical outcomes. Furthermore, patients generally prefer to remain in a familiar environment. The medical support of home care services honors that preference. Home care includes the use of medical equipment, telemedicine monitoring, and portable diagnostic tools. The technology-intensive services range from the simple recording of vital parameters to the more complex management of the entire therapeutic path.

In this context, healthcare related to remote clinical decision-making can be considered a critical operation. Situation Awareness (SA) deals with this decision process to maintain and understand what is happening in a certain situation and leverage this information to avoid or mitigate eventual risks. Today, the Internet of Things (IoT) has evolved due to an advanced interconnectivity of hardware devices equipped with sensors and actuators. With the advent of IoT in healthcare, many low-cost devices are used to monitor patient’s health status remotely. IoT has been widely adopted for both in-home and in-hospital care. Therefore, it is possible to achieve a situation-aware IoT smart home/health environment by: (i) provisioning sensed data to enable monitoring of these environments, (ii) detecting situations based on recorded event logs, and (iii) trigger an action based on the recognized situations. Specifically, a situation-aware system in eHealth gives the opportunity to personalize the therapeutic path for every patient, considering the biological characteristics of the pathology, the aspects of the clinical history, and the living environment.

The complexity and rise of data in the healthcare sector mean that Artificial Intelligence (AI) will increasingly be applied within the field, having the potential to transform many aspects of patient care, as well as administrative processes. Inherently, specific AI sub-fields such as machine learning (ML), deep learning (DL), natural language processing (NLP) and process mining enable healthcare stakeholders and medical professionals to identify healthcare needs and solutions faster and more accurately, using computational models to make informed medical or business decisions quickly. In particular, process mining for healthcare is an appropriate method for extracting information from event logs that are scattered throughout the health system and for defining workflows to be analyzed.

In more recent times, the healthcare industry may be revolutionized by edge-enabled AI, where a series of embedded sensors and IoT devices interconnect to provide diverse intelligent services for the well-being of in-home patients. As Edge AI technology continues to mature, it is increasingly being included in healthcare decision making, as AI is now a key use case for edge computing and edge is a significant enabler for AI.

Metabolic syndrome is a group of risk factors that include abdominal fat, high blood pressure, high blood sugar, and unhealthy cholesterol levels. Treatment focuses on addressing each of these conditions. The goal is to reduce the chances of blood vessel disease, heart disease and diabetes. In most cases, the best treatment for metabolic syndrome rests with the patient. Changes in a patient’s behavior, such as eating healthier and exercising more, are the first things the doctor will suggest. By adopting some healthy habits, the patient may...
be able to completely eliminate risk factors. Quality healthcare outcomes depend upon patients’ adherence to recommended treatment regimens. Patient non-adherence can be a pervasive threat to health and well-being and carry an appreciable economic burden as well.

Thus, the focus of the present paper is on how to measure adherence to therapy of in-home care patients. Therefore, this work proposes an innovative method to measure adherence to therapy aimed at providing awareness of the patient’s current situation in healthcare environments, specifically in-home care. Given the difficulty of quantitatively assessing a patient’s behavioral rigor in following prescriptions, the idea is to exploit process mining techniques at the Edge in a smart home environment when a patient has to stick to therapy, in order to define the level to which the patient’s actions (such as drug intake, adherence to diets, physical activity) are in line with the physician’s indications for the management of the metabolic syndrome.

The paper is organized as follows. Section II provides an overview of related work and technologies which were investigated as background knowledge, including fundamental notions of metabolic syndrome and process mining. Section III shows our proposal on a real-world dataset, taking into account process mining techniques to model the medical prescription as a process model and test the adherence to the therapy of patients with conformance checking. Section IV outlines the experimental setting and assesses event log data with state of the art evaluation metrics. Finally, Section V presents the concluding remarks.

II. BACKGROUND AND RELATED WORK

A. Metabolic Syndrome

The metabolic syndrome is a markedly heterogeneous nosological and clinical entity, still in the process of being defined, represented by the simultaneous association of alterations such as obesity, impaired glucose tolerance, dyslipidemia, and arterial hypertension. In Italy, the prevalence of the metabolic syndrome is around 20% [1]. Progressive increases in the occurrence of metabolic syndrome are age-related, and the prevalence of insulin resistance and glucose intolerance usually increases with increasing age.

Generally, metabolic syndrome is diagnosed if at least three of the symptoms listed in Table I are present.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waist circumference (cm)</td>
<td>≥ 102 men ≥ 88 women</td>
</tr>
<tr>
<td>Fasting Blood Glucose (mg/dL)</td>
<td>≥ 100</td>
</tr>
<tr>
<td>Blood Pressure (mmHg)</td>
<td>≥ 130/85</td>
</tr>
<tr>
<td>Fasting Triglycerides (mg/dL)</td>
<td>≥ 150</td>
</tr>
<tr>
<td>Cholesterol HDL (mg/dL)</td>
<td>&lt; 40 men &lt; 50 women</td>
</tr>
</tbody>
</table>

The primary goal of therapy for metabolic syndrome is to improve or normalize reduced insulin sensitivity. The initial therapeutic approach to metabolic syndrome may be an attempt to abolish its initial causes, e.g. atherogenic diet, sedentary lifestyle, overweight, and obesity. Therapeutic strategies include pharmacological interventions and supplemental home treatments. After diagnosing metabolic syndrome or any of its components, making healthy lifestyle changes can help prevent or delay serious health problems, such as heart attack or stroke. The recommended treatment with home remedies for the management of the metabolic syndrome, validated by an Italian physician, is summarized in the Table II.

<table>
<thead>
<tr>
<th>Event</th>
<th>Prescription</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before Breakfast</td>
<td>- glucose measurement</td>
</tr>
<tr>
<td></td>
<td>- weight measurement</td>
</tr>
<tr>
<td></td>
<td>- diabetes pill</td>
</tr>
<tr>
<td>After Breakfast</td>
<td>- pressure pill</td>
</tr>
<tr>
<td>Mid-Morning</td>
<td>- pressure measurement</td>
</tr>
<tr>
<td>Before Lunch</td>
<td>- glucose measurement</td>
</tr>
<tr>
<td></td>
<td>- diabetes pill</td>
</tr>
<tr>
<td>Afternoon</td>
<td>- exercise: walking, cylette or tapis rulant</td>
</tr>
<tr>
<td>Before Dinner</td>
<td>- glucose measurement</td>
</tr>
<tr>
<td></td>
<td>- pressure measurement</td>
</tr>
<tr>
<td></td>
<td>- diabetes pill</td>
</tr>
<tr>
<td>After Dinner</td>
<td>- pressure pill</td>
</tr>
</tbody>
</table>

Therefore, the treatment described in the above table may be understood as a process model to be carefully followed. Deviations from such a model constitute a level of adherence to therapy that can be difficult to assess. Understanding the degree of adherence to a therapy and indicating an overall percentage of compliance can be a very effective tool for both the patient and the doctor, in order to be able to quickly intervene to help when this degree of adherence falls below a certain threshold.

B. Process Mining

The increasing adoption of Hospital Information Systems (HISs) and Electronic Health Records (EHRs), together with the recent IoT advancements, are allowing smart homes and hospitals to measure a variety of patient- and process related data. Specifically, process mining has emerged as a suitable approach to analyze, discover, improve and manage real-life and complex processes, by extracting knowledge from event logs. In particular, healthcare processes are renowned as complex, flexible, multidisciplinary and ad-hoc, and, thus, they are difficult to manage and analyze with traditional model-driven techniques [3].

Process mining techniques have been used for various use cases within the healthcare domain. Examples include discovering the actual order of activities in a patient’s treatment trajectory [4], evaluating the extent to which clinical guidelines have been followed [5], and predicting patient outcome based on how the treatment is performed [6].

Starting point for process mining is an event log. All process mining techniques assume that it is possible to sequentially record events such that each event refers to an activity (i.e.,
a well-defined step in some process) and is related to a particular case (i.e., a specific execution of activities in a determined order). Case traces are lists of events associated to steps. A workflow (or process model) is therefore a formal specification of how an activity sequence can be composed and can end in a valid process. A process consists of a suitable combination of different tasks performed by agents. A task is a generic piece of work to be executed. In particular, an existing process model can be compared with an event log of the same process. Conformance checking is a specific process mining technique that can be used to check if reality, as recorded in the log, conforms to the model and vice versa. Hence, conformance checking techniques need an event log and a model as input. The output consists of diagnostic information showing differences and commonalities between model and log. In other words, it evaluates how well an event log that records the actual executions matches the model.

Petri net is widely used in process mining as a basis for validation. It can be described as a graphical method of the formal definition of logical interaction between components or the flow of activities in a complex system. Petri net is particularly well suited for modelling concurrency and conflict, sequencing, conditional branch and looping, synchronization, limited resource allocation, and mutual exclusion. The log of a series of events will refer as a task. When the task is executed a token is produced at the start state, then when an event is executed if an edge with the same name is enabled (that is to say there is a token on the preceding state) then the token is consumed and one produced at each connected state. The fact that an edge may lead to more than one state allows for parallel execution. If there is no enabled edge matching the event then one which has the same name is chosen at random and the same process is followed, without the consumption.

C. Related Work

AI-based approaches represent powerful tools as decision support systems in Healthcare (see [7]–[11] for an overview). Also, IoT is emerging in healthcare industry to face several future challenges such as data privacy, security issue, real-time processing, low power consumption, and need for medical expertise [12], [13]. A recent work [3] provided the most recent overview of literature and the extensive number of reviewed papers regarding process mining in healthcare. Works in [14]–[17] focused on an Ambient Assisted Living scenario, in which an intelligent home environment is created to adaptively assist users at home. The underlying rationale assumes an on-edge system, connected to one or more wearable medical devices, that is able to collect, analyze and interpret real-time clinical parameters and to provide an EWS-like clinical risk measurement. The latter works focused on an evolutionary behavior by dividing the learning problem in two simpler ones, in order to correctly distinguish between low-urgency and emergency scenarios, with the possibility of selecting the most convenient configuration able to choose the most appropriate classifier even when the feature set does not allow a robust model selection.

Authors, in [18], presented a situation-aware Internet of Things (IoT) service coordination platform based on the event-driven Service-oriented architecture (SOA) paradigm. The results showed that the IoT situation-aware service coordination platform worked well. A service coordination behaviour model was used, and a Publish/Subscribe based unified message space was used to facilitate asynchronous asynchronous communication and on-demand distribution of sensory data in a large-scale distributed IoT environment. In the work [19], authors presented a situation-aware dynamic IoT services coordination approach. Focusing on the definition of formal situation event pattern with event selection and consumption strategy, an automaton-based situational event detection algorithm is proposed. Lopes et al. [20] proposed a software architecture aimed to provide awareness of the patient’s current situation in healthcare environments, specifically in Intensive Care Units.

The authors [21] proposed an architecture based on an IoT-Edge architecture able to detect deviations from a patient’s care path. In this perspective, Ardito et al. [22] provided a formalization of the clinical guidelines management method. Through the application of this, it is evident how patient monitoring is increasingly improved.

Based on Edge Computing and AI techniques, works in [23], [24] presented a formalization of an intelligent software module aimed at checking the adherence to the care path assigned to a patient at home being remotely monitored. The need for more healthcare choices for Situation Awareness technologies is reflected in the pursuit of established practices related to Telemedicine, allowing teleconsultations with specialists and a more flexible monitoring of the patients at home.

III. MATERIALS AND METHODS

A. Data Collection

Real-world data, containing the logs of the behavior of patients with metabolic syndrome, within the project of Italian Ministry of Education, University and Research MIUR “Progetto Cluster Tecnologici Nazionali - Tecnologie per gli Ambienti di Vita: Active Ageing At Home” (AAAH)\(^1\).

The dataset recorded information of patients during tele-monitoring at home, collected through IoT and medical devices, and describes various activities, such as meals, drug intake, physical exercises, measurement of vital signs, weight measurement. The data have been collected from a cohort of 19 continuously monitored patients with common characteristics located in the Italian Apulia region. The data collection process was carried out over 30 days, sampling all the actions performed by a patient during the daily routine from a series of medical devices which convey sensed data to an edge module in charge of gathering and supplying them to an event log in a standard format compatible with process mining techniques.

\(^1\)Avviso per lo sviluppo e il potenziamento di cluster tecnologici nazionali. Area TAV: Tecnologie per gli Ambienti di Vita. Active Ageing At Home Project, PON Code CTN01_00128_297061. Official portal of the project: http://activeageingathome.eresult.it/
### B. Data Preprocessing

For this experimental analysis, the sensed data are stored in chronological order and in several different tables, i.e. a table containing the drugs taken by the patient, a dietary table, a table relating to blood pressure measurements, one relating to the measurements of blood glucose, one relating to body weight measurements, and finally a table with the different exercises performed. The presence of different tables is determined by the fact that the different equipped sensors store their data in the respective tables, given the consistency required for the data collection based on the patient taken into consideration.

An operation of data aggregation from the various tables in the database under consideration has been carried out, in order to create a single dataset in .csv format useful for subsequent processing. Hence, a standardization operation has been carried out. In particular, the timestamp has been standardized to the dd/mm/yyyy hh:mm format. Diet-related activities have been standardized in breakfast, lunch and dinner. Physical activity has been unified with a single exercise activity, and the type of performed activity has been indicated in one categorical value between “walking”, “cyclette”, and “tapis roulant”. The blood glucose measurement activity has been standardized with glucose_measurement as well as the weight, standardized to weight_measurement. Also, the pressure measurement has been standardized with pressure_measurement, reporting the indication of the systolic and diastolic measurement in a “systolic/diastolic” way. While, the following notation has been adopted regarding the patient’s drug intake during the day: pressure_pill and diabetes_pill for the intake of drugs related to blood pressure and diabetes, respectively.

Ultimately, the data were aggregated by patient using a unique id, and subsequently sorted according to the timestamp provided, in such a way as to obtain the sequence of activities performed by the same patient in chronological order during the period of monitoring. In this way the data stored in different tables have become rows of a single dataset with the respective label. Indeed, in Table III is described a fragment of the final event log including for each case: case ID, timestamp, activity, and value.

### C. Modeling the Metabolic Syndrome with Process Mining

The idea is to model the problem of metabolic syndrome in-home treatment as a process model and then feed the retrieved dataset to the process model in order to run the conformance checking technique, achieving, in this way, an awareness of some threat. The event log analysis is performed at the end of the day, in order to assess its compliance along the entire path.

We modeled the ideal process defined in Table II with the 3 Process Mining modeling algorithms, implemented in the Python library PM4Py [25]:

1) **Alpha Miner** [26]: With an event log as the input, the Alpha Miner algorithm derives various relations between the activities occurring in the event log. These relations are used to produce a Petri net that represents the log;

2) **Heuristic Miner** [27]: Heuristics Miner is an algorithm that provides a way to handle with noise and to find common constructs (dependency between two activities). The basic idea is that infrequent paths should not be incorporated into the model. The output of the Heuristics Miner is an object that contains the activities and the relationships between them, that can be then converted into a Petri net;

3) **Inductive Miner** [28]: The basic idea of Inductive Miner is to find a prominent split in the event log (there are different types of splits: sequential, parallel, concurrent, and loop). After finding the split, the algorithm recurs on the sub-logs (found by applying the split) until a base case is identified. Inductive miner can discover robust process models from noisy and incomplete data, and can produce a Petri net.

Process models modeled using Petri nets have a well-defined semantic: a process execution starts from the places included in the initial marking and finishes at the places included in the final marking.

In Figure 1 is depicted the Petri net of the process model for metabolic syndrome management. Specifically, the figure describes the daily routine that a patient has to follow according to the prescription. The Petri net represented by Heuristic Miner algorithm contains hidden transitions and has the advantage of having different parameters that can be used for the elimination of unimportant clusters. The most important ones are the dependency threshold, with a default value of
Fig. 1. Petri net related to the metabolic syndrome in-home prescription designed with the Heuristic Miner algorithm.

IV. EXPERIMENTS

Conformance checking is a technique to compare a process model with an event log of the same process. The goal is to check if the event log conforms to the model, and, vice versa. In PM4Py, two fundamental techniques are implemented: token-based replay and alignments.

Token-based replay is a heuristic technique, which uses four counters (produced tokens, consumed tokens, missing tokens, and remaining tokens) to compute the fitness of an observation trace based on a given process model [29]. A trace is fitting according to the model if, during its execution, the transitions can be fired without the need to insert any missing token. If the reaching of the final marking is imposed, then a trace is fitting if it reaches the final marking without any missing or remaining tokens.

Alignment-based replay is a technique, which performs an exhaustive search to find out the optimal alignment between the observed trace and the process model. Hence, it is guaranteed to return the closest model run in comparison to the trace [30].

It is possible to compare the behavior contained in the log and the behavior contained in the model, in order to see if and how they match. Four different dimensions exist in process mining, including replay fitness, precision, generalization, and simplicity.

- **Replay Fitness**: The calculation of the replay fitness aim to calculate how much of the behavior in the log is admitted by the process model. For token-based replay, the percentage of traces that are completely fit is returned, along with a fitness value that is calculated [31]. For alignments, the percentage of traces that are completely fit is returned, along with a fitness value that is calculated as the average of the fitness values of the single traces.

- **Precision**: The different prefixes of the log are replayed (whether possible) on the model. At the reached marking, the set of transitions that are enabled in the process model is compared with the set of activities that follow the prefix. The more the sets are different, the more the precision value is low. The more the sets are similar, the more the precision value is high. This works only if the replay of the prefix on the process model works: if the replay does not produce a result, the prefix is not considered for the computation of precision. Hence, the precision calculated on top of unfit processes is not really meaningful. There exist two approaches for the measurement of precision: ETConformance (using token-based replay) [32], and Align-ETConformance (using alignments) [33].

- **Generalization**: A model is general whether the elements of the model are visited enough often during a replay operation (of the log on the model). A model may be perfectly fitting the log and perfectly precise. Hence, to
measure generalization a token-based replay operation is performed, and the generalization is calculated [34].

- **Simplicity**: A dimension that evaluates how simple the process model is to understand for a human. It is defined taking into account only the Petri net model; the criteria used for simplicity is the inverse arc degree [35].

### A. Evaluation

For a better understanding of the evaluation phase, the results, for each of the three process mining algorithms, have been reported on the histogram in Figure 2.

In our investigation, it is crucial that a patient follows the doctor’s prescription perfectly and should not reverse, introduce or forget activities. A small note can be made about certain activities, such as those relating to measuring blood pressure or blood sugar as if they were reversed they would not involve anything serious, compared to forgetting to take a drug, for example. This means that the evaluation metrics we want to be maximized are precision and replay fitness.

In particular, these two metrics behave differently depending on the inspected process mining algorithms. For a deeper understanding of how much patients have been adherent to the medical prescription for metabolic syndrome treatment, it is in general evident how the replay fitness addresses this compliance checking. Replay fitness is above 79% in all process mining modeling algorithms, suggesting that most of the patients observed the model behavior correctly. Precision, instead, is quite swinging, showing that, although the fitness reaches good results for all the algorithms, Alpha Miner and Inductive Miner algorithms show opposite behaviors. This means that they do not rightly and fairly represent the actual behavior of patients. In fact, they are not able to capture some wrong activities, such as revers actions in the medical prescription, even though those actions are included in the medical prescription. A good trade-off between replay fitness and precision is given by the Heuristic Miner. It gives a more accurate representation of the patients’ actual behavior. Selecting the Heuristic Miner as best model representation for conformance checking, it is possible to note that the 95% of patients followed the medical prescription entirely, while the 75% of them followed the medical prescription precisely.

The generalization metric aims to maximize model-supported behaviors that are not part of the system and are not present in the event log. This, in our case, is not good as the patient must follow the activities present in the medical prescription, without introducing new activities unless recommended by the doctor. As shown in Figure 2, the generalization values are typically low, below the 17% for all process mining algorithms, witnessing that model does not generalize. This is good news for our scenario, as it means that patients do not perform activities as different as those prescribed by the doctor.

Simplicity assumes less importance as it is not linked to the behavior of the process model obtained but to its simplicity in terms of network and therefore of understanding. The value of simplicity for all algorithms used on the medical prescription event log is a high indication that the model processed by the three algorithms is quite simple.

### V. Concluding Remarks

The advances in cognitive IoT edge compute systems trigger the use of AI solutions to achieve awareness in healthcare. Process mining in eHealth is an emerging research field with sound future scope to provide a situation awareness computational intelligent model to measure adherence to therapy.

In this paper, we presented an operational approach to assessing patients’ behavioral rigor in following medical prescriptions, specifically for metabolic syndrome management. By modeling the medical prescription as a process model, the conformance checking techniques from the process mining field have been exploited to show several adherence evaluation indexes tested with real-world event logs from in-home patients with metabolic syndrome. Analytical experiments showed how adherent they were to the therapy and, interestingly, such data helped analysts in finding the most suitable process model representation algorithm that best fitted this problem.

As future considerations, a systematic approach could be proposed to transform patient activity and information from clinical data into situations. By doing so, we would be able to predict the future impact so that we could finally adapt their behavior according to the users’ intentions.

### ACKNOWLEDGMENT

The authors would like to thank Michele Stasolla and Giulio Mallardi for their help in running the experiments. This work was funded by the Italian P.O. Puglia FESR 2014 – 2020 (project code 6ESURE5) ‘SECURE SAFE APULIA’.

### REFERENCES

