Extracting Models of Dynamic Human Health Behaviour from Real-Life Personal Health Data

Afonso Sarmento(✉) and José Machado

LASI/ALGORITMI, University of Minho, Braga, Portugal
pg55831@uminho.pt

Abstract. The rise of lifestyle-related chronic diseases has increased the need for data-driven methods that capture how health behaviours unfold in daily life. Mobile and wearable technologies now enable continuous, multimodal health data collection, but most behavioural models use static or aggregated representations that overlook temporal dependencies and context.

In this work, we adopt a process-oriented perspective, modelling health behaviour as a dynamic process and present a novel methodological contribution: a pipeline that transforms multimodal streams (meals, physical activity, continuous glucose monitoring (CGM), self-reports) into event logs suitable for both traditional and object-centric process discovery. We apply and compare several process discovery algorithms to a longitudinal multimodal dataset from individuals with diabetes, deriving models that capture interactions among activities, physiological measures, and daily routines. Our results indicate that Object-Centric Process Mining (OCPM) can extract interpretable, clinically meaningful insights from heterogeneous sensor and self-report data, supporting its use as a methodological foundation for modelling dynamic health behaviour in real-world care.

Keywords: Process Mining · Human Health Behaviour · mHealth

1 Introduction

According to the World Health Organization (WHO), the global prevalence of diseases aggravated by unhealthy lifestyle, habits and behaviours, such as cardiovascular diseases, cancer or diabetes, is increasing. This highlights the necessity of developing strategies of facilitating the practice of health-promoting behaviours in individuals [1]. This includes engagement in self-monitoring behaviours and the adoption of self-care practices that may improve long-term health outcomes and overall well-being. These advances have been supported by technologies such as wearables or mobile apps (especially mHealth apps) with the purpose of helping this self-control [2,3].

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2027
O. Gervasi et al. (Eds.): ICCSA 2026 Workshops, LNCS 16761, pp. 601–618, 2027.
https://doi.org/10.1007/978-3-032-30527-5_39

Recent work has shown that human health behaviour follows a process-like structure [4]. A process-oriented perspective is particularly relevant because it preserves the temporal ordering, timing, and interaction of behaviours, enabling the identification of behavioural sequences that precede health outcomes and can be targeted by personalised and temporally optimised interventions [5]. A dynamic perspective analyses behaviour not as a fixed attribute but as a time-dependent process shaped by variability that emerges across multiple levels of human activity [6].

Process mining differs from traditional data mining by explicitly reconstructing how behaviours unfold in time rather than extracting aggregated patterns from static features. By reconstructing event order and revealing branching points and deviations, process mining produces an interpretable, process-centred view that is directly useful for identifying actionable behavioural sequences and designing timely, personalized interventions [7].

There is a lack of approaches capable of accurately interpreting the human health behaviour dynamics from these data. This represents a clear gap between the availability of data and the tools needed to model health behaviour as a dynamic process. As a result, it remains unclear how different process discovery techniques model health behaviour, what behavioural characteristics they capture or miss, and which modelling choices lead to interpretable and clinically plausible representations.

This study addresses the following research question: How can we model health behaviour processes from real-life health data?

To address this gap, we propose a process-oriented approach based on process mining techniques, guided by three sub-questions: Which process-discovery and object-centric techniques exist for modelling health behaviour, and how do they compare?; How can multimodal sensor and self-report data be transformed into event logs for process discovery?; Which discovery algorithms and modelling choices produce the most interpretable and robust models on real-world diabetes data?

We aim to transform heterogeneous sensor and self-report records into event logs suitable for process discovery, compare discovery algorithms in terms of interpretability and robustness, and validate the discovered models using process-mining quality metrics and comparison with established reference patterns. The central contribution of this work is the development and evaluation of an end-to-end methodological pipeline that operationalizes OCPM for multimodal, real-life personal health data. While the paper includes a comparative benchmark of algorithms, its primary novelty lies in the transformation of continuous sensor streams into object-centric event logs, enabling the capture of complex interactions between physiological responses and daily behaviours that traditional case-centric approaches overlook.

2 State of the Art

2.1 Foundations of Human Health Behaviour

Health behaviour refers to the set of patterns, activities, and habits undertaken with the purpose of prevention and health maintenance, improvement and to promote well-being [8].

Human behaviour is central to health and illness. Behaviours, activities, and self-regulation shape the biopsychosocial state, including motivations, actions, and processes that support well-being, such as the practice of regular physical activity and a balanced diet [8]. Health-promoting behaviours are associated with lower morbidity and mortality and better quality of life [9].

Because individuals vary in capability, context, and motivation, personalised behavioural models can tailor interventions to their specific needs, increasing effectiveness and improving health outcomes [10,11].

In this context, numerous conceptual theories and models (e.g. COM-B, Theory of Planned behaviour, HAPA, Social Cognitive Theory) have been developed to explain the behaviours. While these models provide valuable frameworks for understanding health behaviour and behaviour change, they have key limitations for digital health personalisation [10]. Although some models allow for change across stages or motivational states, they still lack formal mechanisms for stochasticity, multidimensional interactions, partial structure, and fine-grained context adaptation.

To address this gap, this study specifies explicit characteristics of human behaviour, detailed in the next section. We operationalize each characteristic for digital health data by representing observations as time-stamped events and objects, encoding context as event attributes and temporal windows, and allowing probabilistic transition models and optional paths in discovered processes.

Health Behaviour Characteristics The characteristics that guide this study are the following:

- **Partially structured or unstructured** - Behavioural sciences describe human behaviour as a combination of habitual routines and adaptive responses to changing contexts, aligning with the assumption that human behaviour does not always follow rigid, linear, or formally defined rules [12]. It is important to consider that human behaviour can also be structured. Daily and weekly patterns, habitual activities or the circadian cycle exhibit structured and predictable patterns. This supports the use of discovery techniques that allow for flexible and non-linear process representations instead of rigid or fully structured models [13].
- **Stochastic** - The individual responses vary due to internal states, environmental influences and unobserved factors [14]. As a result, behaviour cannot be assumed to be fully predictable, even under similar circumstances. This assumption motivates the use of probabilistic and entropy-aware process discovery techniques, which can capture behavioural variability rather than enforcing deterministic models [12].

- **Multidimensional** - Human behaviour emerges from interactions across multiple domains, being consistent with behavioural theories that emphasize cognitive, emotional, and social determinants [15]. This supports the adoption of modelling approaches capable of integrating and jointly analysing multiple behavioural dimensions, enabling the discovery of interdependencies across heterogeneous data sources. [12].
- **Context-dependent** - The behaviour varies according to situational, cultural, social, and environmental conditions and, in different contexts, the same individual may behave differently. Probabilities, context-specific entropy, and interaction measures can quantify the extent to which behaviours shift in response to context changes. This motivates the use of approaches that can incorporate this information into behavioural modelling, enabling the analysis of how behavioural patterns adapt across varying situational and environmental conditions. [12].

Collectively, these characteristics establish a robust foundation for the selection and evaluation of process mining techniques that can adequately capture the dynamic, variable, and context-sensitive characteristics of human health behaviour.

2.2 Data-Driven Approaches

Data-driven methodologies are being developed as means to enable personalisation and variability within digital health interventions and tools. These methods rely on event-based data that capture user behaviours and contexts, providing the empirical traces needed to operationalize personalised interventions. These approaches aim to identify patterns and relationships in data, enabling the construction of models that capture how individuals behave in practice rather than how behaviour is theoretically described [7].

Although these techniques can uncover associations between variables, most data mining techniques are not process-oriented. Without this temporal structure, these patterns remain invisible, the sequences cannot be reconstructed, and process-oriented patterns are overlooked. [7].

In the context of mHealth, personalisation depends not only on which behaviours occur, but also on when and in what order they occur, as interventions must respond to individuals' evolving behavioural trajectories [16]. For clarity, in a control-flow perspective, a process is modelled as ordered sequences of events (traces) governed by control-flow constructs such as sequence, choice, concurrency and loops. By aggregating and transforming these behavioural data into static features, these temporal dynamics are ignored, limiting the ability to reconstruct behavioural processes and constraining the design of personalised and adaptive digital interventions.

This deeper understanding motivates the use of process-oriented, data-driven techniques to model and analyse real-world human health behaviour, which is essential to develop real-time and personalised digital health interventions.

Process Perspective: Process Mining According to Wil van der Aalst, process mining is between machine learning and data mining on the one hand, and process modelling and analysis on the other [14]. The idea of process mining is to discover, monitor and improve real processes by extracting knowledge from event logs readily available in today's systems [14].

Process mining offers approaches explicitly designed to improve process analysis. One of its main advantages is its objectivity, as it relies on real data event logs [17]. Each event is associated with an event type, which specifies the lifecycle transition or state change it represented by the event within a process execution. In other words, it provides semantic information about how an event relates to the progression of an activity or process instance [17].

Process mining techniques reconstruct process models, enabling identification of temporal structures such as routines, transitions, deviations, and behavioural cycles [18]. By preserving the order and timing of events, process mining captures aspects of human behaviour that are lost in traditional data mining—such as the sequence in which actions occur, the repetition of behavioural loops, or the conditions under which decisions are made [18]. Consequently, process mining provides dynamic representations and has the potential to support more precise and adaptive personalisation in digital interventions. A key challenge of process mining lies not only in appropriate data use, but also in selecting modelling formalisms capable of representing essential properties of human behaviour. Although good data use can return powerful conclusions, poor data preprocessing and use can lead to unexpected results. Even with high-quality data, models that fail to capture behavioural concurrency, stochasticity, or cross-object relations will limit the validity and usefulness of the resulting insights [7].

Object-Centric Process Mining Traditional process mining was created with the capability of dealing with a single type of objects at a time, where each event corresponds to exactly one object and the process models created through them describe their patterns and lifecycle. While this assumption is suitable for well-structured organizational processes, it becomes restrictive when modelling complex, multidimensional phenomena such as human health behaviour, where multiple interacting entities evolve simultaneously, such as physiological measurements, activities or context factors.

Object-centric appears with the goal of working without the classical constraints, with multidimensionality in the type of objects [19]. An object is an identifiable entity in the system that participates in events and around which processes are observed, analysed, and modelled [20]. Object-centric has the advantages of providing a multidimensional representation of the processes, by explicitly modelling objects and their relationships, which in traditional process mining is abstracted away.

This capability is essential for capturing the complexity of health behaviour and enables process-oriented analyses from multiple perspectives without requiring data re-extraction or the definition of a case notion [19].

The next step is to examine the process discovery techniques that are suitable for this case study.

3 Process Discovery Techniques

Process mining aims to improve operational processes through the use of event data during the discovery of process models from event logs [14,21].

Its importance lies in its potential to identify processes, standardize clinical practices, support evidence-based decisions and improve the quality of healthcare and health habits [22]. These processes consist of sequences of daily activities and health-related events, such as meals, physical activity, and physiological responses, with the goals of improving patient outcomes and monitoring patient trajectories. They involve clinical and non-clinical activities performed by diverse resources and can vary widely. Personal health management processes are highly dynamic and individual-centred, which makes their analysis and optimisation particularly challenging [22].

Over the past years, numerous process mining techniques have been developed, as summarized in Table 1.

Table 1. Comparison of explored process mining techniques and their supported characteristics.

			Behavioural				Technical							
Name	Reference	Year	Multidimensional	Stochastic	Partially structured	Context-dependent	Control-flow	Loops	Duplicate tasks	Hidden tasks	Noise	Missing data	Clustering-based	Algorithmic approach
Markov	[23]	1998	x	x	x		x	x		x	x	x		x
Alpha/Alpha+	[23]	2004					x							x
Heuristics Miner	[23,24]	2006	x				x	x		x	x	x		x
Decision Miner	[25]	2006				x								x
Genetic Miner	[23]	2007					x	x		x	x	x		
DT Genetic Miner	[23]	2007					x	x	x	x	x	x	x	
Fuzzy Miner	[23]	2007	x		x		x	x			x	x	x	x
Alpha++	[23]	2007	x				x							x
ILP Miner	[23]	2009					x			x		x		x
FSM Miner/Petrify	[25]	2010					x	x	x	x				x
FSM Miner/Genet	[23]	2010					x	x						x
Inductive Miner	[25]	2013					x	x			x		x	x
Inductive Miner (infrequent)	[25]	2014					x	x		x	x			x
Fodina	[12]	2017	x		x		x	x	x	x	x	x		x
Split Miner	[12]	2017					x	x			x			x
Hybrid ILP Miner	[25]	2018					x	x		x	x			x

While control-flow models provide an effective representation of the structural ordering of activities, standard process discovery algorithms generally fail to capture the stochastic variability that is intrinsic to health behaviour data, including probabilistic transition choices, temporal variability, and the influence of unobserved events. In typical applications, sequence discovery often begins by extracting transitions as ordered pairs of consecutive events from the event log. The aggregation of these pairs produces a transition matrix and an associated directed graph that summarises the most frequent paths and branching patterns. However, they do not encode explicit temporal information or probabilistic dynamics. Consequently, we adopt control-flow models as an interpretable

structural core and explicitly augment them with temporal and pattern-based analyses to characterise inter-event intervals, transition probabilities, and the impact of missing or latent events [5].

4 Methods

4.1 Dataset

To analyse process mining techniques in health data, this study is guided by dataset-related selection criteria that define the data's scope and selection criteria, ensuring it serves as a good use case and satisfies the following conditions:

- **Be an open-source dataset**, allowing transparency, reproducibility of results and facilitating comparative and follow-up studies.
- **Include sensor data**, preferably from two different types of sensors or modalities.
- **Reflect real-world conditions**, meaning data were collected without instructing participants to change their usual routines or behaviours.
- **Contain manual logs**, corresponding to self-reported activity log data.
- **Meet a minimum continuous recording period**, sufficient to capture longitudinal patterns.

With all these requirements in mind, the *Longitudinal Multimodal Dataset of Type 1 Diabetes* was identified as a suitable choice.

This dataset was created by collecting health data from 17 individuals, with ages between 23 and 70 years [15]. Data registers were collected during a continuous 90-day period [15].

Health-related information was captured through dedicated platforms using standardized protocols. CGM data were derived from multiple commercially available systems (LibreView, Dexcom, Medtronic), whereas insulin administration data were obtained from insulin pump platforms or electronic logs documenting multiple daily injection regimens. Dietary intake was recorded using Nutritics software, and physical activity and sleep were quantified with Garmin Forerunner 45 smartwatches, providing high-resolution, sensor-based measurements [15].

4.2 Framework

Traditionally, process mining is designed and studied following specific guidelines for its execution: the Process Mining Project Methodology (PM^2). These guides are created with the aim of improving process performance and compliance to rules and regulations, as represented in Fig. 1 [26].

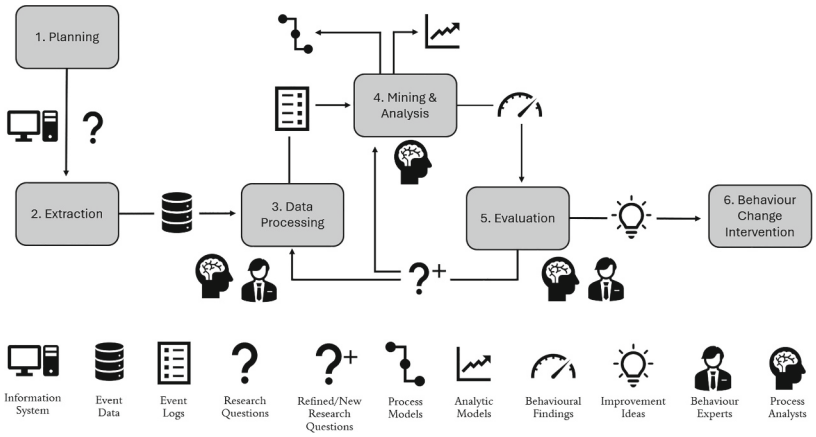

Fig. 1. Overview of the PM^2 methodology.

Event Log Creation The creation of the event logs is the first part in the processing and study of process mining techniques. It is through event logs that processes are analysed and the information, patterns and inferences are executed [7].

For these experiments, we test the case IDs of *meals* and *days*, with and without filters for events of hyperglycaemia and hypoglycaemia. The study of different case notions provides a more accurate representation of real-world processes. The four slicing strategies we used were the Day strategy, which groups all events for a patient within the same calendar day, Day_HH which subdivides each day into shorter time blocks to capture finer temporal granularity, the Meal strategy which defines cases around recorded meal timestamps, and Meal_HH which extends meal-based cases by applying a fixed temporal window to include surrounding events. Specifically, Day_HH subdivides each day into four fixed 6-h blocks (00 h–06 h, 06 h–12 h, 12 h–18 h, 18 h–00h). Meal_HH extends the case duration to a window starting 30 min before the meal and ending 4 h after, to capture the full postprandial glucose curve and late insulin effects.

We implement a compact, rule-based inference pipeline to map continuous sensor data into discrete event types. First, signals are preprocessed (time alignment across sensors, smoothing using a 15-min moving average to reduce high-frequency sensor noise, and outlier removal defined as the exclusion of glucose rate-of-change values exceeding 5 mg/dL/min, which are physiologically implausible). Then the following behavioural events are inferred programmatically:

- **Meal start**, derived from recorded meal timestamps.
- **Bolus insulin**, detected from timestamped insulin-administration records and associated to the nearest meal within a predefined window.
- **Postprandial peak**, identified as the local maximum within a three-hour post-meal window that exceeds the preprandial baseline (the preprandial

baseline is computed as the median glucose level during the 30 min preceding the meal timestamp).
- **Recovery (return to baseline)**, defined as the first timestamp after the peak at which glucose returns to or falls below the preprandial baseline.

To illustrate the practical application of the proposed pipeline, a representative behavioural vignette is defined as follows: a 'Meal' event (e.g., breakfast) is recorded alongside a synchronous 'Insulin Bolus' administration. These events trigger a temporal window that monitors the CGM stream, identifying a 'Post-Prandial Blood Glucose (PPBG)' episode. This episode is characterized by a transition from a 'Baseline' state to 'Hyperglycemia', followed by a 'Recovery' event once the glucose levels return to the target range.

Pipeline For this pipeline, a set of process discovery algorithms was selected to be tested: Heuristics Miner, Inductive Miner, Integer Linear Programming (ILP) Miner, Alpha Plus and Fodina, because together they align with the behavioural characteristics outlined above.

The object-centric Heuristics Miner extends frequency-based dependency analysis to object-centric event logs by computing directly following relations conditioned on object types and instances. The object-centric Inductive Miner adapts recursive block-structured process discovery to object-centric event logs by identifying control-flow constructs while respecting objectâĂŞevent associations. It produces sound, structured process models that support conformance checking and reproducible behavioural analysis, making it suitable for hypothesis-driven evaluation and clinical validation.

Following the PM^2 methodology, the pipeline executes the data extraction, transformation, discovery, and evaluation phases. The event data is loaded and preprocessed by renaming the case identifier, activity label, and timestamp columns, normalizing timestamps, and ordering events chronologically. During the transformation phase, two event log representations are constructed: a case-centric event log exported in eXtensible Event Stream (XES) format, and an Object-Centric Event Log (OCEL), where relevant object types and their associations with events are explicitly defined. This object-centric transformation enables events to be linked to multiple object instances and supports the application of object-centric discovery techniques.

The object-centric event log is structured around PPBG episodes, which serve as the primary object type for the analysis. The OCEL was constructed by mapping events to objects using a 'many-to-many' relationship: a single physical activity event could be linked to two different 'PPBG Episode' objects if it occurred within their respective 4-hour post-meal windows. Quality measures for OCPM were computed using the OC-Discovery framework, where fitness and precision are calculated relative to object-event associations rather than linear traces. In this representation, each event can reference multiple object instances, allowing behavioural and physiological entities to coexist within the same event context. Each episode represents a temporally delimited physiological response following a meal and links behavioural and physiological events such as meals,

insulin administrations, glucose peaks, and recovery phases. Using PPBG as the central object enables the representation of interactions between behavioural actions and physiological responses within the same episode while preserving temporal ordering and relationships between entities.

From standardized data event log, we perform process discovery by applying the data to the different established algorithms described in Sect. 3. Then, we conduct a quantitative evaluation across multiple dimensions, as referred in Sect. 4.3. Conformance checking via token-based replay (TBR) is then executed by simulating each trace against the discovered Petri Net. This allows us to identify fitting and non-fitting cases and to assess how well the observed behaviour is captured by the model and, consequently, by its Petri Net representation. All discovery experiments were executed using the same preprocessing pipeline and identical parameter settings to ensure comparability between techniques. Figure 2 provides an overview of the proposed pipeline.

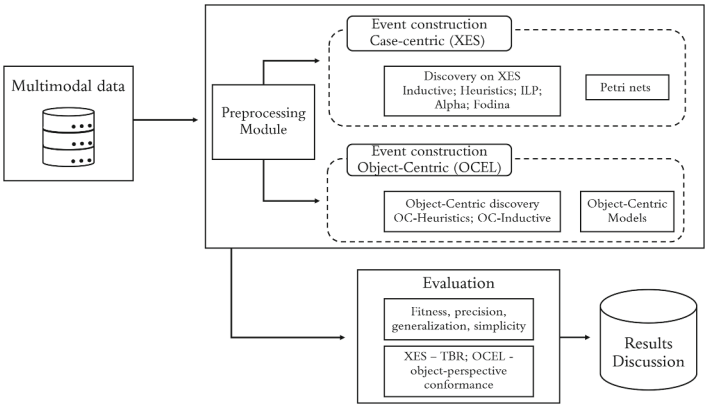

Fig. 2. Overview of the PM^2-aligned pipeline, illustrating the parallel case-centric (XES) and object-centric (OCEL) discovery paths and their evaluation.

4.3 Evaluation Parameters

To analyse the tested models, we consider a set of evaluation dimensions that are used as indicators to evaluate the quality of the executed models. These dimensions, shown on Fig. 3, tell us how well a technique can create a reliable process model, capable of capturing process behaviour with the greatest resemblance to what reality is. These evaluation dimensions also indicate how well the techniques can capture behavioural variability inherent in human-driven processes, including optional, repeated, or noisy activities.

To evaluate a discovery model, two main dimensions are considered: fitness and comprehensibility.

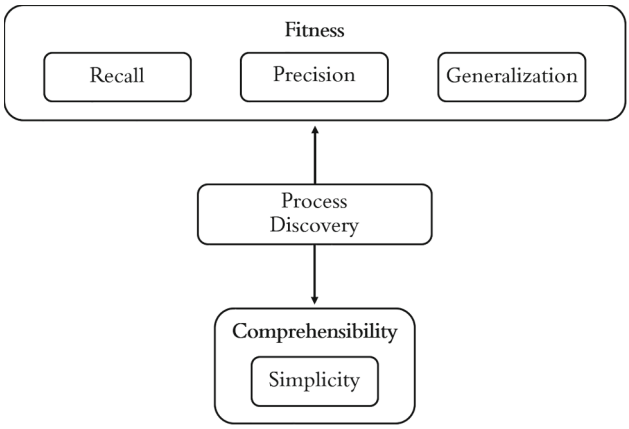

Fig. 3. Framework of evaluation metrics for assessing process discovery techniques.

Four dimensions are evaluated: fitness, precision, generalisation, and simplicity. Fitness, operationalised as trace fitness, measures the extent to which the behaviour recorded in the event log can be replayed by the discovered model, and thus captures the recall of observed traces. Precision checks that the model neither overfits the data nor becomes too general by allowing behaviour not supported by the log. Generalisation assesses whether the model correctly represents not only observed behaviour but also unseen yet plausible behaviour that may occur in the real process [23].

Comprehensibility includes simplicity and evaluates how easily humans can understand and interpret a discovered process model, considering its complexity and ease of interpretation. It is largely determined by structural complexity, including size, structure, and concurrency [23].

Together, these metrics provide a comprehensive evaluation of model quality by considering behavioural fidelity, precision over unobserved behaviour, generalisation capacity, and structural interpretability.

5 Results

This section reports the obtained results and evaluation about all the tested models, following the evaluation criteria defined in Sect. 4.3. Highlight to fitness, precision, generalisation and simplicity. The results are explicit in Table 2.

Inductive Miner and ILP Miner achieved the highest fitness values (1.00 and 0.9985), indicating a strong ability to reproduce the event-log behaviour. Inductive Miner's perfect fitness suggests an exact replay of the observed traces, while ILP Miner combines similarly high fitness with very low precision and low simplicity.

Regarding precision, Fodina and Alpha Plus reached the highest values (1.000 and 0.9183), suggesting better control over unobserved behaviour. Despite this,

Table 2. Evaluation of process discovery techniques across fitness, precision, generalisation and simplicity.

Technique	Precision	Fitness	Generalisation	Simplicity	CS
Heuristics Miner	0.5064	0.9608	0.9290	0.4959	0.7230
Inductive Miner	0.4388	1.0000	0.9455	0.6570	0.7603
ILP Miner	0.3679	0.9985	0.9774	0.1648	0.6272
Alpha Plus	0.9183	0.8063	0.9797	0.8036	0.8770
Fodina	1.0000	0.4297	0.9823	0.3861	0.6995

Fodina's precision coincides with lower fitness (0.4297), indicating that it may fail to reproduce a substantial portion of the observed traces, effectively capturing a limited subset with high exactness.

For generalisation, Fodina (0.9823) and Alpha Plus (0.9797) scored highest. Alpha Plus combines high generalisation with high precision and moderate fitness, indicating a balanced ability to go beyond the observed traces while maintaining behavioural control. In contrast, ILP Miner shows very high fitness and generalisation but lower precision and simplicity, typical of models that fully reproduce the log.

The combined quality metric (CS), computed as an equally weighted average of fitness, precision, generalisation and simplicity (0.25 each), ranks Alpha Plus highest overall, as it offers the most balanced trade-off, followed by Inductive Miner and Heuristics Miner. Alpha Plus does not achieve the best fitness individually, showing that the combined metric rewards balanced performance and better captures heterogeneous behavioural patterns than extreme values on a single dimension.

In OCPM, Table 3 shows that both Heuristics Miner and Inductive Miner achieve perfect fitness, which means they fully replay the recorded object-centric traces. However, Inductive Miner has higher precision and generalisation than Heuristics Miner, as well as greater simplicity, resulting in a better combined quality score.

Table 3. Object-centric process discovery techniques across the four evaluation metrics.

Technique	Precision	Generalisation	Simplicity	Fitness	CS
Heuristics Miner	0.6103	0.9657	0.6577	1.0000	0.8084
Inductive Miner	0.9972	0.9774	0.7333	1.0000	0.9270

Together, these results suggest that, using the object-centric interpretation of the dataset, Inductive Miner produces models that are both truthful to the log and more constrained with respect to unobserved behaviour, while still permitting good generalisation over the observed traces. By contrast, Heuristics Miner

is more open to behaviour not present in the log, permitted by lower precision, which may reflect either loose control in the object-centric models or sensitivity to concurrency/noise introduced by multiple object types.

To assess the robustness of the techniques and the possible effect of different case definitions, an experimental matrix was constructed using multiple case IDs derived from the same dataset (daily and meal-based event logs, with and without filters for hypoglycaemia and hyperglycaemia), as seen in Table 4.

Table 4. Experimental matrix with combined score values across different case attribute definitions.

	Day	Meal	Meal_HH	Day_HH
Heuristics Miner	0.6596	0.8067	0.8132	0.6125
Inductive Miner	0.6892	0.8989	0.7905	0.6657
ILP Miner	0.5822	0.7085	0.6506	0.5709
Alpha Plus	0.9800	0.8744	0.5587	0.5322
Fodina	0.7616	0.7035	0.4270	0.6810

Results show that Heuristics and Inductive Miner remain relatively stable across the different case definitions, although some variability is observed, particularly for Inductive Miner. In contrast, Alpha Plus and Fodina exhibit higher sensitivity to case-definition granularity, with substantial performance degradation under more granular and dense case notions. ILP Miner consistently produces lower combined scores and shows little variation between case definitions, indicating uniformly weaker performance rather than strong sensitivity to changes in case granularity.

These results suggest that case definition plays a critical role in process discovery from health data, and that not all algorithms are equally resilient to changes in temporal or semantic aggregation.

The two discovered Petri Nets exhibit different structural characteristics that reflect the underlying modelling choices.

The object-centric Petri Net (Fig. 4) is substantially more compact and more clearly organised. By representing post-prandial episodes and related entities as distinct object types, the object-centric miner separates concerns that would, otherwise, be blended in a single case stream. This separation creates a net in which relations between events are more explicit and semantically coherent. Recurring episode patterns (for example, the sequence meal → peak → recovery cycles) are represented as compact looping structures that are easy to follow visually and to interpret analytically.

By contrast, the traditional Petri Net produced using meal-based case identification (Fig. 5) is considerably more complex. Aggregating heterogeneous attributes and concurrent behaviours into a single case identifier forces the discovery algorithm to encode a large number of distinct behaviours and contextual

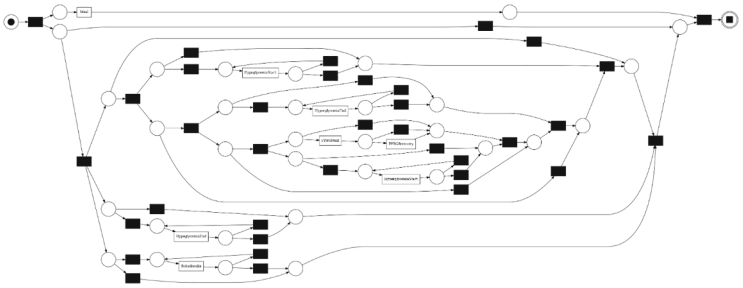

Fig. 4. Petri Net obtained using the Object-Centric Heuristics Miner.

arrangements within one net, resulting in many places, transitions, and alternative paths. This structural complexity difficult human interpretability and, as observed in the literature, tends to reduce model generalisation: highly complex networks can reproduce the observed traces closely (high fitness)but often exhibit behaviour that overfits the sample and performs worse on plausible unseen traces.

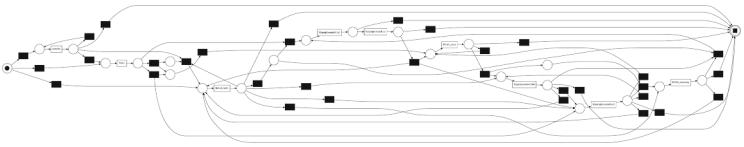

Fig. 5. Petri Net discovered using the traditional Heuristics Miner with meal-based case identification, filtered for hypoglycaemia and hyperglycaemia events.

Taken together, these observations suggest a pragmatic trade-off for multimodal health data: the object-centric representation improves interpretability by making inter-entity relations explicit and by exposing looping and concurrent patterns in a semantically meaningful way, while case-centric flattening into a single case stream can generate more complex nets that attempt to accommodate many attributes simultaneously.

6 Discussion

The results address the three sub-questions defined in the introduction. Regarding the first sub-question, Table 1 provides a structured comparison of available process discovery and object-centric techniques, and Tables 2, 3 and 4 quantify their performance across four quality dimensions on real health data. Regarding the second sub-question, the event log construction pipeline described in Sect. 4.2 demonstrates that multimodal sensor and self-report streams can be systematically transformed into both XES and OCEL formats through rule-based inference and temporal windowing. Regarding the third sub-question, Heuristics

Miner and Inductive Miner consistently achieve the best balance of fitness, interpretability, and robustness across case definitions, with object-centric variants further improving precision and generalisation.

The main goal of this study is to investigate how human health behaviour can be modelled from real-world, multimodal data and to evaluate whether process mining can produce meaningful, personalized representations of such behaviours.

The results show that Heuristics Miner presents a profile characterized by high fitness and generalisation, but with moderate precision. This behaviour suggests that the discovered models reproduce a large portion of observed traces while still allowing some additional variability beyond the log.

In the context of human health data, which are usually affected by noise and large variability, this kind of behaviour can be beneficial. The technique reliably represents recurring patterns without over-constraining the model so much that slightly atypical traces are excluded. For clinical use, however, moderate precision means that the models may include false behaviour.

Inductive Miner exhibits a performance profile of achieving perfect fitness, high simplicity and high generalisation, while maintaining stability across different case definitions. The lower precision observed is consistent with the algorithm's abstraction mechanism, which compresses observed behaviour into a smaller structural model, improving interpretability but sometimes omitting fine-grained differences between traces.

In this domain, model interpretability and structural clarity are essential for clinical validation, analysis, and communication with non-technical stakeholders. Although Inductive Miner exhibits lower precision on this dataset, this behaviour can be attributed to its structural assumptions and abstraction mechanisms, which compress observed behaviour into compact representations. Consequently, some fine-grained behavioural aspects may be lost.

Alpha Plus, despite achieving the highest combined score, exemplifies the limitations of trusting only on aggregate metrics. This observation reinforces the importance of aligning the evaluation criteria with domain-specific objectives rather than optimizing generic quality scores.

The object-centric analysis reinforces these conclusions by revealing structural and behavioural properties that are difficult to capture under traditional case-centric. In contrast to single-case modelling, OCPM enable a more truthful representation of multimodal and context-dependent human health behaviour, where actions emerge from the interaction of behavioural and context objects rather than from a single linear case. Variability across days, meals, and context conditions are not treated as deviation from a single trace, but as a natural consequence of interacting behavioural dimensions.

Additionally, the object-centric representation tends to reduce overgeneralisation and over-fitting effects observed in the case-centric setting. By preserving object relations, concurrency, and shared events, object-centric models provide a more grounded behavioural space, offering greater expressive power while remaining suitable for abstraction, comparison, and conformance analysis.

This study has some limitations. The analysis is based on a single, diabetes-focused dataset, which limits the generalisability of the findings. Given the small sample size ($n = 17$) and the specific nature of Type 1 Diabetes management, the behavioral patterns discovered should be viewed as a methodological validation of the OCPM pipeline rather than universal claims about human health behavior. Second, the discovered models and their quality assessments are sensitive to methodological choices, particularly case-ID definitions, temporal segmentation strategies, and preprocessing decisions. Additionally, the modelling approach adopted in this study is predominantly deterministic, relying on control-flow representations that encode observed behavioural sequences as fixed structural relations. Human health behaviour, however, is stochastic, context-dependent, and probabilistic in nature. As a consequence, deterministic process models cannot fully capture the intrinsic variability, uncertainty, and adaptive dynamics that characterise real-world behaviour. While the discovered models approximate dominant behavioural patterns, they cannot represent human behaviour with total exactness.

Together, the findings suggest that no single technique universally captures all relevant aspects of human health behaviour. Instead, Heuristics Miner and Inductive Miner emerge as suitable for this domain due to their complementary strengths: the capacity to capture frequent patterns with high fidelity, while Inductive Miner provides structurally sound and interpretable models that support formal analysis and clinical validation. Their relative robustness to case definitions further supports their applicability in real-world health datasets, where temporal aggregation and semantic boundaries are not clearly defined.

6.1 Future Work

Future work can focus on extending both the methodological and application-oriented aspects of this work. From a methodological perspective, research can explore how the development or adaptation of algorithms that more closely align with specific behavioural patterns of interest, particularly in scenarios where existing approaches fail to adequately capture such behaviours' dynamics. An important direction consists in designing an interface capable of assisting users by providing contextually relevant and actionable feedback, while simultaneously promoting greater user autonomy. Such an interface would support personalised monitoring, behaviour recommendations and adaptive guidance, providing informed decision-making and progressive behavioural adjustment over time.

7 Conclusions

The findings provide insights about how such methods can support a deeper understanding of behavioural dynamics while informing the design of more personalised and adaptive digital health solutions.

Overall, the results demonstrate that no single process discovery technique can be assumed as ideal for modelling human health behaviour.

OCPM emerges not as a technical alternative, but as a more appropriate modelling interpretation for health-related data, as it enables the representation of interacting behavioural and context entities. Techniques such as Heuristics Miner and Inductive Miner, particularly under object-centric interpretations, demonstrate characteristics that align more closely with the requirements of human behaviour modelling, offering a balanced trade-off between behavioural fidelity, interpretability, and generalisation.

References

1. World Health Organization. Noncommunicable diseases (2021). https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases
2. Topol, E.: The Creative Destruction of Medicine: How the Digital Revolution Will Create Better Health Care. Basic Books (2019)
3. Almeida, M., Machado, J., Sousa, R., Peixoto, H.: AI-powered automation in healthcare: a multi-agent approach with LLM. Commun. Comput. Info. Sci. **2644**, pp. 132–143. Springer, Deutschland, Journal Abbreviation: Commun. Comput. Inf. Sci (2025)
4. Brzychczy, E., Aleknonytė-Resch, M., Janssen, D., Koschmider, A.: Process mining on sensor data: a review of related works. Knowl. Inf. Syst. **67**(6), 4915–4948 (2025)
5. Ferreira, D.R.: Control-flow perspective. In: Ferreira, D.R. (ed.) A Primer on Process Mining: Practical Skills with Python and Graphviz, pp. 15–29. Springer, Cham (2020)
6. White, D.R.: Dynamics of human behavior. World Futures **65**(1), 25–35 (2009)
7. Saylam, R., Sahingoz, O.K.: Process mining in business process management: concepts and challenges. In: 2013 International Conference on Electronics, Computer and Computation (ICECCO), pp. 131–134 (2013)
8. Conner, M.: Health behaviors. In: International Encyclopedia of the Social & Behavioral Sciences (2015)
9. Sztyler, T., Völker, J., Carmona, J., Meier, O., Stuckenschmidt, H.: Discovery of personal processes from labeled sensor data: an application of process mining to personalized health care. In: Business Process Management Workshops. LNBIP, vol. 256, pp. 164–175. Springer, Cham (2015)
10. Cancela, J., Charlafti, I., Colloud, S., Wu, C.: Digital health in the era of personalized healthcare: opportunities and challenges for bringing research and patient care to a new level. In: Digit. Health, 7–31. Elsevier (2021)
11. GCruz, ., Guimarães, T., Santos, M.F., Machado, J.: Decentralize healthcare marketplace. Procedia Comput. Sci. **231**, 439–444. Elsevier B.V. (2024)
12. Rehse, J.-R., Leemans, S.J.J., Fettke, P., van der Werf, J.M.E.M.: On process discovery experimentation: addressing the need for research methodology in process discovery. ACM Trans. Softw. Eng. Methodol. **34**(1), 3:1–3:29 (2024)
13. Zee, P.C., Attarian, H., Videnovic, A.: Circadian rhythm abnormalities. Continuum: Lifelong Learn. Neurol. **19**(1 Sleep Disorders), 132–147 (2013)
14. van der Aalst, W.: Process Mining. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-662-49851-4
15. Alsuhaymi, A., et al.: A Longitudinal multimodal dataset of type 1 diabetes. Sci. Data **12**(1), 1379 (2025)

16. Fallon, M., Rehse, J.-R., Heinzl, A.: Human behavior mining: a framework for theorizing about mHealth behavior using digital trace data. J. Assoc. Inf. Syst. **26**(3), 633–667 (2025)
17. Leemans, M.: Hierarchical Process Mining for Scalable Software Analysis. Technical report, Eindhoven University of Technology (2018)
18. Conradi, R., Jaccheri, M.L.: Process modelling languages. In: Derniame, J.-C., Kaba, B.A., Wastell, D. (eds.) Software Process: Principles, Methodology, and Technology. LNCS, vol. 1500, pp. 27–52. Springer, Heidelberg (1999). https://doi.org/10.1007/3-540-49205-4_3
19. van der Aalst, W.M.P.: Object-centric process mining: an introduction. In: Cerone, A. (ed.) ICTAC 2021. LNCS, pp. 73–105. Springer, Cham (2023)
20. Celonis. Getting started with object-centric process mining. Accessed 12 Mar 2026
21. van der Aalst, W.M.P., Carmona, J. (eds.): Process Mining Handbook. LNBIP, vol. 448. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-08848-3
22. Aversano, L., Iammarino, M., Madau, A., Pirlo, G., Semeraro, G.: Process mining applications in healthcare: a systematic literature review. PeerJ Comput. Sci. **11**, e2613 (2025). Publisher: PeerJ Inc
23. De Weerdt, J., De Backer, M., Vanthienen, J., Baesens, B.: A multi-dimensional quality assessment of state-of-the-art process discovery algorithms using real-life event logs. Inf. Syst. **37**(7), 654–676 (2012)
24. Kurniati, A.P., Kusuma, G.P., Wisudiawan, G.A.A.: Implementing heuristic miner for different types of event logs. ARPN J. Eng. Appl. Sci. **11**(8), 5114–5120 (2016)
25. Folz-Weinstein, S., Bergenthum, R., Desel, J., Kovář, J.: ILP2 miner - process discovery for partially ordered event logs using integer linear programming. In: Gomes, L., Lorenz, R. (eds.) PETRI NETS 2023. LNCS, vol. 13929, pp. 59–76. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-33620-1_4
26. van Eck, M.L., Lu, X., Leemans, S.J.J., van der Aalst, W.M.P.: PM^2: a process mining project methodology. In: Zdravkovic, J., Kirikova, M., Johannesson, P. (eds.) CAiSE 2015. LNCS, vol. 9097, pp. 297–313. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-19069-3_19