International Journal of Medical Informatics 220 (2026) 106601

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf

Original Research Article

Discovering heart failure patient pathways through event enrichment and abstraction

Anton Antonov [a,d,*], Harry H. Beyel [a], Marlo Verket [b], Christopher T. Schwanen [a], Viki Peeva [a], Marco Pegoraro [a], Julia Brandts [b,c], Dirk Müller-Wieland [b], Nikolaus Marx [b], Wil M.P. van der Aalst [a,d], Katharina Marx-Schütt [b]

[a] Chair of Process and Data Science (PADS), RWTH Aachen University, Ahornstraße 55, Aachen 52074, Germany
[b] Department of Internal Medicine I, RWTH Aachen University Hospital, Pauwelsstraße 30, Aachen 52074, Germany
[c] Imperial Centre for Cardiovascular Disease Prevention, School of Public Health, Imperial College London, UK
[d] Data Science & Artificial Intelligence, Fraunhofer FIT, Birlinghoven Castle, Sankt Augustin 53757, Germany

ARTICLE INFO

Keywords:
Process Mining
Healthcare
Patient Pathway Discovery
Clinical Data Integration
Event Abstraction
Patient Stratification

ABSTRACT

Background: Integration of electronic health data from patients with heart failure into process mining could provide insights to the patient journey. However, the current approaches have challenges in addressing all aspects of patient care.

Methods: A retrospective cohort study of 234 patients with heart failure who were admitted to an outpatient heart failure clinic was conducted. We developed a two-stage pipeline to integrate multidimensional data aspects, such as patient-reported outcome measures, biomarkers, medication changes and reasons for hospitalization. The pipeline consists of: (1) Data Enrichment, where we derive indicators from disparate sources like outpatient visits and self-reported data and we structure this information into an event log; and (2) Supervised Event Abstraction where low-level events are translated into clinically meaningful concepts using a knowledge graph. We demonstrate the practical integration of clinical event enrichment and knowledge-graph-based abstraction for exploratory heart failure pathway analysis.

Results: We identified four clusters based on patients' backgrounds. Group A included older, multi-morbid patients with advanced HF and more frequent HF hospitalizations and renal-function worsening, while Groups B, C, and D represented ischemic, arrhythmia-associated, and younger lower-comorbidity profiles with simpler care pathways.

Conclusions: These findings demonstrate that this process mining approach offers a practical framework to better understand specific patient pathways and identify mitigation strategies for adverse disease trajectories and mortality. It allows for a more granular understanding of patient journeys and the identification of at-risk cohorts by integrating clinical and patient-reported outcomes. This is a significant step towards more data-driven, patient-centered healthcare.

1. Introduction

Process mining [1] uses event data to model and improve operational workflows. Driven by the widespread adoption of *health information systems* (HISs), its purpose in healthcare is to analyze and optimize administrative processes and, more recently, patient trajectories [2]. Healthcare process mining has largely focused on process discovery [3], i.e., techniques for inferring process models from event data. Nevertheless, its practical application is often hampered by two key challenges.

The first is the high variability in patient journeys. Because medical care is tailored to individual needs, mapping patient trajectories often

* Corresponding author at: Chair of Process and Data Science (PADS), RWTH Aachen University, Ahornstraße 55, Aachen 52074, Germany.
E-mail addresses: anton.antonov@rwth-aachen.de (A. Antonov), beyel@pads.rwth-aachen.de (H.H. Beyel), mverket@ukaachen.de (M. Verket), schwanen@pads.rwth-aachen.de (C.T. Schwanen), peeva@pads.rwth-aachen.de (V. Peeva), pegoraro@pads.rwth-aachen.de (M. Pegoraro), jbrandts@ukaachen.de (J. Brandts), dirmueller@ukaachen.de (D. Müller-Wieland), nmarx@ukaachen.de (N. Marx), wvdaalst@pads.rwth-aachen.de (W.M.P. van der Aalst), kschuett@ukaachen.de (K. Marx-Schütt).

https://doi.org/10.1016/j.ijmedinf.2026.106601
Received 23 January 2026; Received in revised form 3 July 2026; Accepted 7 July 2026
Available online 8 July 2026

Patient ID	Date of Visit	Prescription		Reported Symptoms		Conducted Tests	
		Beta-Blockers	…	Edema	…	NT-proBNP	…
3C3D4	23.01.22	Carvedilol (3x6.25mg)		Never		350 pg/mL	
3C3D4	25.02.22	Carvedilol (3x6.25mg)		Never		375 pg/mL	
3C3D4	15.05.22	Bisoprolol (1x2.5mg)		Sometimes		1375 pg/mL	
…							

(a) An illustrative example of outpatient visit data.

Event ID	Case ID	Time	Activity
1	3C3D4	23.01.22	BB prescribed
2	3C3D4	15.05.22	Worsening Edema
3	3C3D4	15.05.22	BB drug switch
4	3C3D4	15.05.22	BB dose increased
5	3C3D4	15.05.22	NT-proBNP increased

(b) The resulting event log after applying the data transformation steps described in Section 2.3.1.

Fig. 1. Data Enrichment stage applied to the toy outpatient visit data.

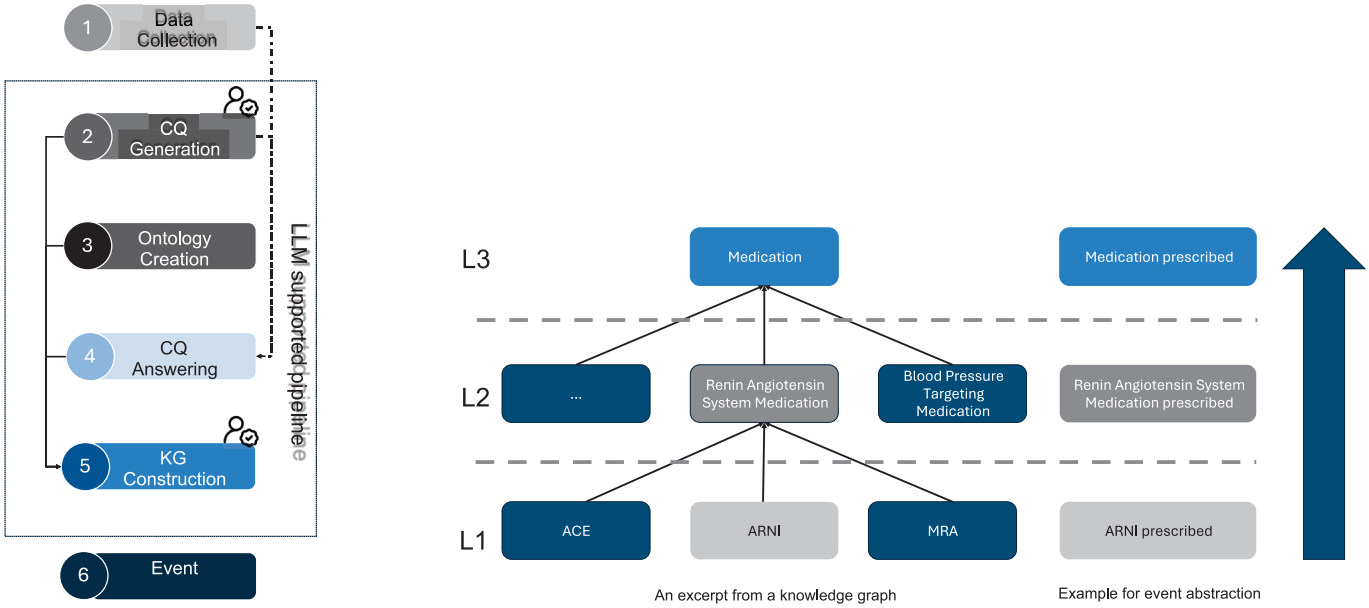

Fig. 2. Event Abstraction Pipeline (a). On the left. Overview of the method for event abstraction. As evident, steps 2–4 are automated using LLMs. However, domain experts review (and if needed, modify) the resulting Competency Questions (CQs) and the knowledge graph. (b) On the right. An example for a knowledge graph and the resulting abstraction for a given event with activity *ARNI prescribed.*

results in overly complex, unreadable "spaghetti models". This problem is especially relevant in analyzing *heart failure* (HF) pathways, where patients exhibit highly diverse backgrounds, comorbidities, and evolving symptom profiles [4]. Illustrating this complexity, a recent study identified approximately 30,000 unique disease trajectories among cardiovascular patients [5], making traditional process discovery algorithms impractical for deriving clear clinical insights.

The second challenge is data integration and the lack of clinical

Table 1
Baseline Characteristics of the Patient Cohort (N = 234).

Characteristic	Value
Demographics	
Female Sex, n (%)	50 (21.4)
Age, mean ± SD (years)	65.3 ± 12.9
LVEF, mean ± SD (%)	37.88 ± 9.96
Cardiovascular History and Procedures, n (%)	
Pacemaker/ICD Implantation	104 (44.4)
Myocardial Infarction	94 (40.2)
Percutaneous Coronary Intervention	90 (38.5)
Atrial Fibrillation	76 (32.5)
Coronary Artery Bypass Grafting	52 (22.2)
Peripheral Artery Disease	23 (9.8)
Stroke	17 (7.3)
Other Comorbidities, n (%)	
Hypertension	124 (53.0)
Chronic Kidney Disease	94 (40.2)
Diabetes Mellitus	56 (23.9)
Hypercholesterolemia	53 (22.6)
Chronic Obstructive Pulmonary Disease	34 (14.5)

Abbreviations: ICD, implantable cardioverter-defibrillator; SD, standard deviation; LVEF, Left Ventricular Ejection Fraction

depth. Existing healthcare process mining research has focused on procedural aspects, such as waiting times [6] or high-level clinical outcomes like deaths and hospitalizations [7,8]. While the need for a holistic patient view is widely acknowledged [9], existing methods fall short of capturing longitudinal clinical patterns. For example, analyses may treat biomarker tests as events but ignore their actual values, thereby failing to capture changes in a patient's condition [8]. Others track pattern changes only at an aggregate process level, omitting individual variations [10]. Recent studies targeting HF treatment pathways highlight this limitation. While offering valuable insights, they predominantly restrict their scope to high-level procedural events [11]. Consequently, the resulting process models lack the depth necessary to explain the clinical changes that precede these outcomes, as treatment-related trends are omitted. This leaves a gap for a methodology capable of managing high process variability while integrating diverse clinical aspects into process models.

To address this, we propose a systematic pipeline designed to discover interpretable, clinically meaningful HF patient pathways. Our methodology transforms diverse medical data – including outpatient visits, self-reported metrics, and admission records – into patient-centered event logs. Our method comprises two distinct stages: data enrichment, which integrates clinical aspects into the event data, and event abstraction, which reduces variability by elevating low-level aspects into clinically relevant states. Through this pipeline, we bridge the gap between complex medical data and actionable process mining in the cardiovascular domain.

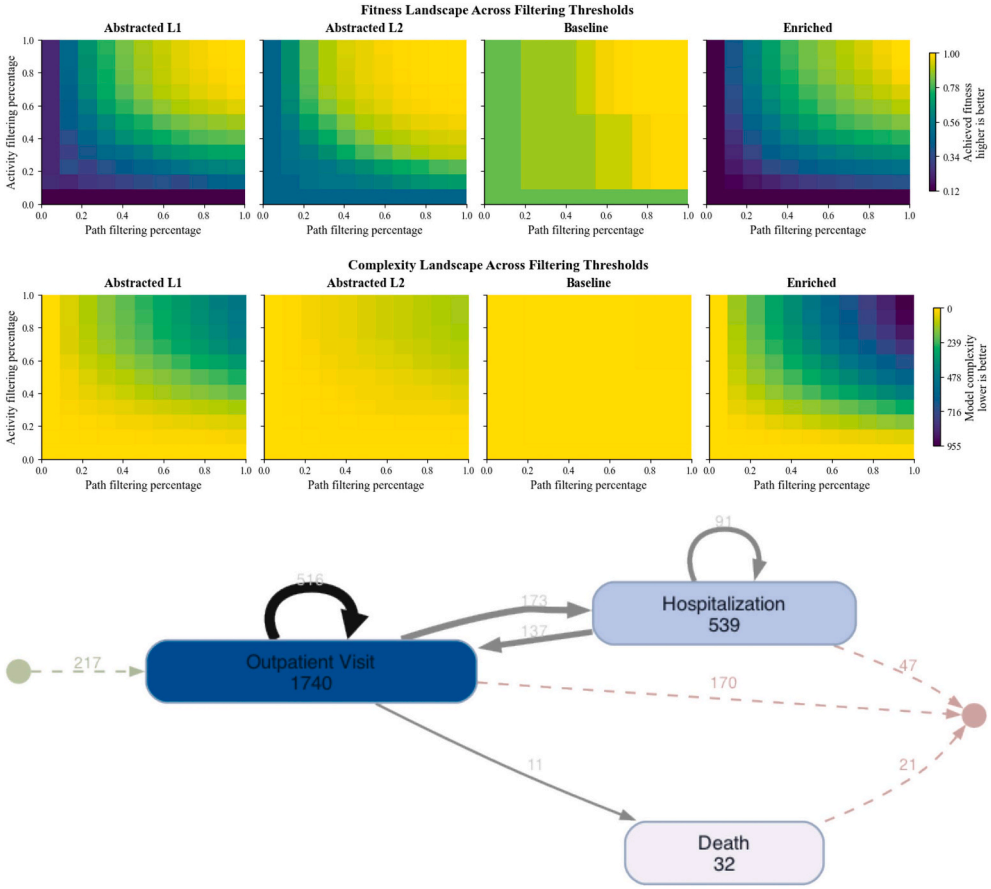

Fig. 3. (a): Fitness across different filtering settings Fig. 3(b): Complexity across different filtering settings Fig. 3(c): Baseline model (fitness: 0.984, complexity: 6) Parameters were set to path filtering = 0.7, activity filtering = 0.7 according to the model selection procedure from Section 2.1. Fig. 3(d): Enriched log (fitness: 0.758, raw complexity: 345); Parameters were set to path filtering = 0.5, activity filtering = 0.6 according to the model selection procedure from Section 2.1. Fig. 3(e): Enriched log (fitness: 0.696, raw complexity: 94) with one level of abstraction (L1); Parameters were set to path filtering = 0.8, activity filtering = 0.3 according to the model selection procedure from Section 2.1. Fig. 3(f): Model for enriched log (fitness: 0.911, raw complexity: 36) with two levels of abstraction (L2); Parameters were set to path filtering = 0.9, activity filtering = 0.3 according to the model selection procedure from Section 2.1.

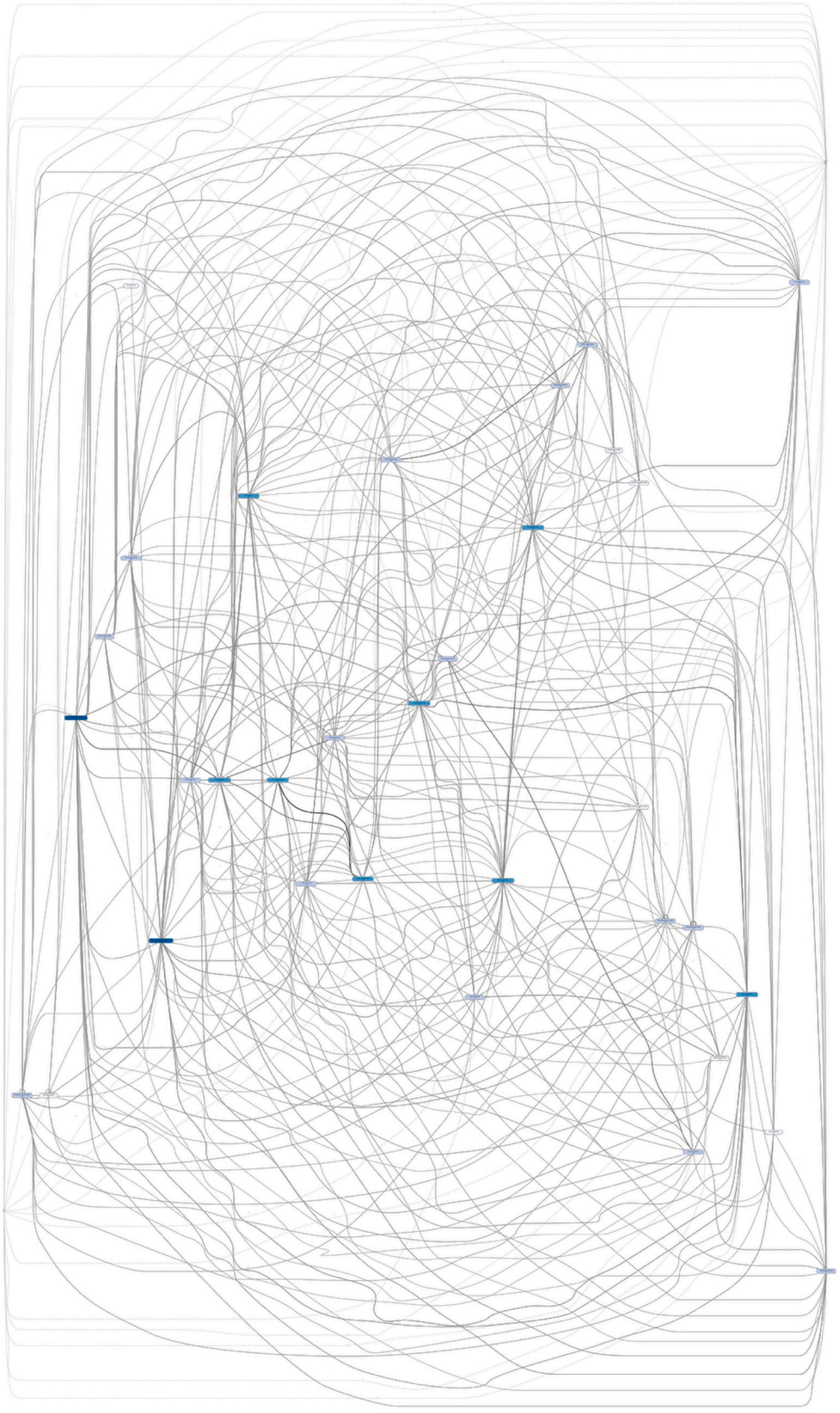

Fig. 3. (*continued*).

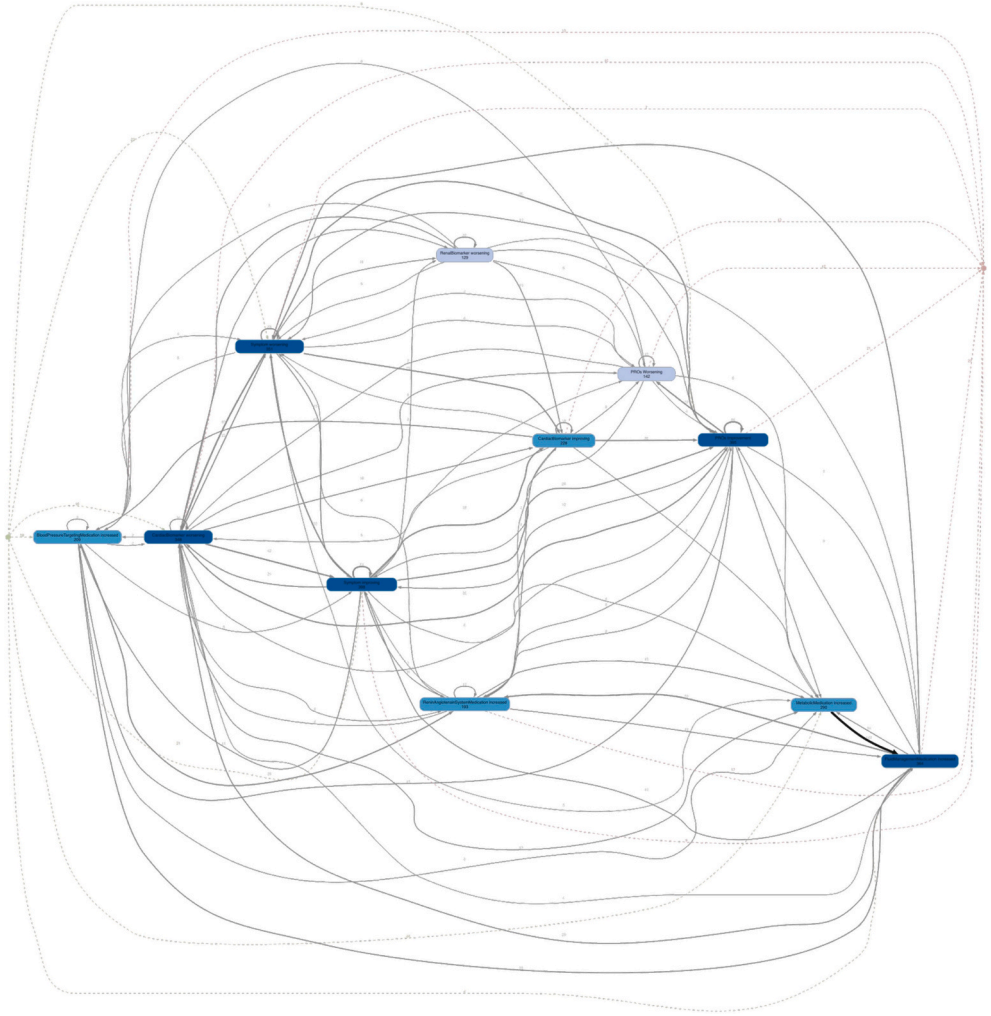

Fig. 3. (*continued*).

2. Methods

2.1. Process mining

Process mining [1] incorporates a family of techniques for reconstructing and analyzing process behavior from event data. In the context of this study, the process we analyze is the longitudinal care trajectories of heart failure patients. The utilized data consists of clinically meaningful events observed during follow-up appointments, e.g., changes in biomarkers, symptoms, medication, patient-reported outcomes, and hospital admissions.

Process mining techniques operate on an event log, i.e., a common representation format of event data. An event log (cf. Fig. 1(b)) contains multiple cases, where, for example, each case corresponds to a patient. Each case consists of a time-ordered sequence of events called a trace. Each event records at least three attributes: the case identifier, in our case, the patient identifier, the timestamp when the event occurred, and the activity. Activities describe what happened, for example "NT-proBNP increased" or "HF hospitalization". Additional attributes, such as biomarker values or demographic information, can be added as event or case attributes.

From an event log, process discovery techniques identify a process model, which summarizes the observed behavior across cases. In this paper, we use directly-follows graphs (DFGs) as the main process representation. In a DFG, nodes represent activities and directed edges denote direct succession between activities. For example, an edge from "SGLT2 prescribed" to "Admission Heart Failure" (cf. Fig. 3(d)) indicates that, for at least one patient, a hospitalization was observed after SGLT2 was prescribed without another recorded event between them. Green dashed edges indicate start activities, while orange dashed edges indicate end activities. DFGs are useful in such settings, as they provide an intuitive representation of the heterogeneous care pathways while avoiding the additional complexity of formal process model semantics. However, when many detailed clinical aspects are included, DFGs can become visually complex and entangled. To keep DFGs interpretable, we applied PM4Py's [12] activity and path filtering settings, which remove less frequent activities and directly-follows relations. Lower thresholds remove more behavior; for example, a path filtering value of 0.1 keeps only the top 10% of directly-follows relations.

As DFGs are similar to transition systems, one can easily measure the model fitness for a given event log [13,14], i.e., by assessing how well the process model describes the underlying behavior. Additionally, as we perform event abstraction, we measure the simplicity of the models using structural complexity, computed as model size [15]. When clustering patients and discussing patient pathways, we selected process models by considering the trade-off between fitness and complexity. We first identified the Pareto-optimal candidate models, defined as models for which no other candidate achieved both equal or higher fitness and equal or lower complexity. For each candidate model i, we computed $d_i = f_i - c_i$, where $f_i \in [0,1]$ denotes model fitness and $c_i \in [0,1]$ denotes the min–max normalized raw complexity. Specifically, the raw complexity value CR_i was normalized as $C_i = \frac{CR_i - min_j CR_j}{max_j CR_j - min_j CR_j}$ across all

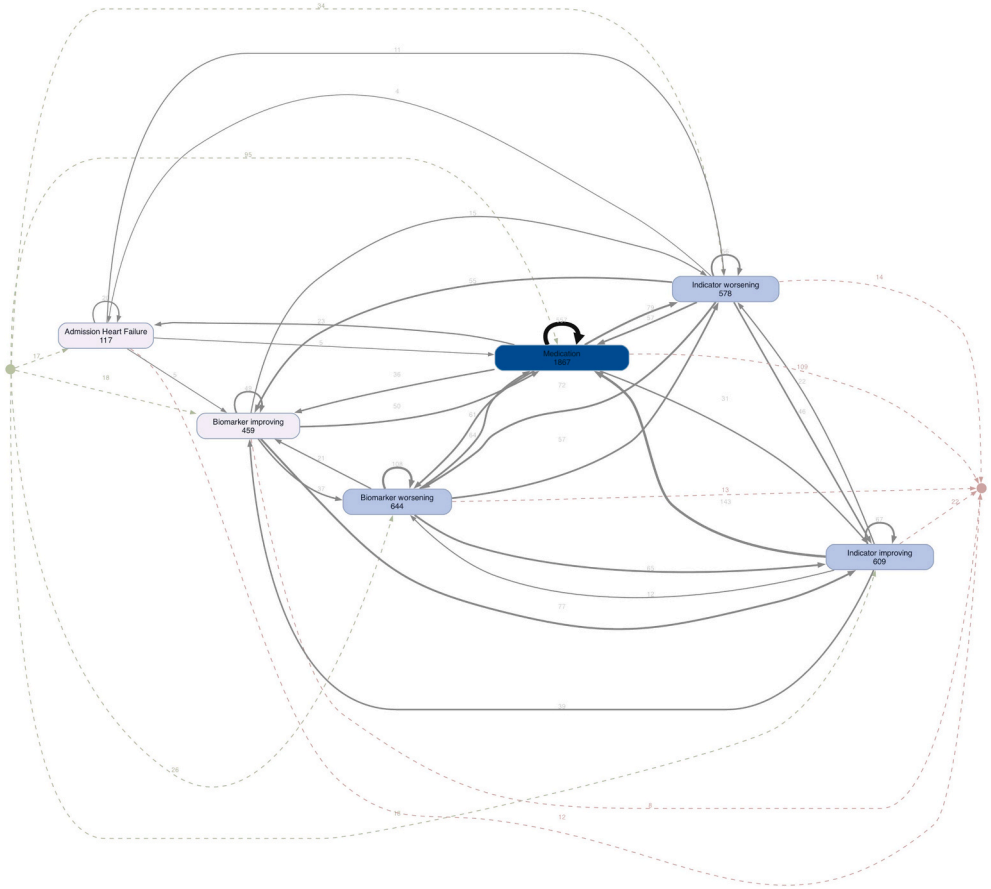

Fig. 3. (*continued*).

models for a given log. Thus, higher d_i values indicate models with higher fitness and lower relative complexity; the ideal model would have $f_i = 1$ and $c_i = 0$.

2.2. Dataset

To evaluate this approach, we used a single-center, prospective, observational cohort study *Aachen Longitudinal Heart Failure and Diabetes Registry Study* (ALIDIA). Eligible participants in ALIDIA are adults ($\geq$ 18 years) with stable chronic heart failure who have visited the outpatient heart failure clinic at Department of Internal Medicine I, RWTH Aachen University Hospital from April 2019 to July 2023. Key exclusion criteria include severe valvular disease, type 1 diabetes, end-stage renal disease, and acute myocarditis. The ALIDIA study is conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the Medical Faculty of RWTH Aachen University (EK 006/19, DRKS00025557). All participants provided written informed consent before inclusion.

The data includes structured electronic health records containing demographic information, diagnoses, symptoms, discharge medication recommendations, patient-reported outcomes (PROs) and laboratory values as obtained from blood tests or electrocardiography in a tabular format and unstructured data as clinical free-text notes in the form of discharge summaries.

2.3. Data enrichment

The data enrichment transforms routinely collected clinical data into a detailed event log.

2.3.1. Changes in biomarkers, measures, symptoms, prescription

Clinically meaningful changes in biomarkers, symptoms, and medication regimens are identified by comparing consecutive patient records.

For biomarkers, events are generated when changes exceed predefined thresholds derived from literature and expert consultation: $\pm$50% in *N-terminal pro b-type natriuretic peptide* (NT-proBNP), $\pm$0.3 mg/L for *cystatin C* (CysC), and 20% for *estimated glomerular filtration rate* (eGFR). Fig. 1 illustrates the resulting event generation process.

Symptoms are recorded as ordinal variables and mapped to numerical values. Changes between consecutive visits are interpreted as symptom improvement or deterioration and recorded as events.

Prescription-related events are extracted from medication records. To ensure comparability across formulations, medications are mapped to standardized therapeutic classes and normalized dosages based on HF treatment guidelines [16], enabling the detection of clinically relevant changes in medication therapy (Fig. 1).

2.3.2. Patient-Reported outcomes

To incorporate the patient's perspective, we included two PRO instruments: the Kansas City Cardiomyopathy Questionnaire (KCCQ) [17] and the Hospital Anxiety and Depression Scale (HADS) [18]. Events are generated when score changes exceed established thresholds for clinical relevance.

2.3.3. Activities from clinical Documents

Clinical reports are transformed into activity labels using LLM-Take [19]. A domain-specific vocabulary covering HF-related procedures and comorbidities is used as a predefined schema for key word extraction. When multiple candidate labels are identified, embeddings generated with BioLORD [20] are used to compute semantic similarity between

Table 2

Characteristic	Group A (N = 82)	Group B (N = 33)	Group C (N = 43)	Group D (N = 76)
Demographics				
Female Sex, n (%)	9 (11.0)	4 (12.1)	10 (23.3)	27 (35.5)
Age, mean ± SD (years)	72.8 ± 9.9	63.6 ± 11.1	70.5 ± 9.6	56.8 ± 12.5
LVEF, mean ± SD (%)	33.8 ± 9.4	37.7 ± 8.3	39.8 ± 9.6	41.2 ± 10.1
Cardiovascular History, n (%)				
Myocardial Infarction	57 (69.5)	27 (81.8)	0 (0.0)	10 (13.2)
Percutaneous Coronary Intervention	57 (69.5)	33 (100.0)	0 (0.0)	0 (0.0)
Coronary Artery Bypass Grafting	37 (45.1)	7 (21.2)	2 (4.7)	6 (7.9)
Peripheral Artery Disease	21 (25.6)	2 (6.1)	0 (0.0)	0 (0.0)
Atrial Fibrillation	32 (39.0)	2 (6.1)	42 (97.7)	0 (0.0)
Stroke	9 (11.0)	0 (0.0)	6 (14.0)	2 (2.6)
Pacemaker/ICD	47 (57.3)	11 (33.3)	17 (39.5)	29 (38.2)
Other Comorbidities, n (%)				
Hypertension	63 (76.8)	12 (36.4)	24 (55.8)	25 (32.9)
Diabetes Mellitus	34 (41.5)	0 (0.0)	12 (27.9)	10 (13.2)
Hypercholesterolemia	22 (26.8)	12 (36.4)	8 (18.6)	11 (14.5)
Chronic Obstructive Pulmonary Disease	24 (29.3)	0 (0.0)	3 (7.0)	7 (9.2)
Chronic Kidney Disease	53 (64.6)	0 (0.0)	23 (53.5)	18 (23.7)

Note: Bold indicates the highest prevalence for each characteristic across the four groups.
Abbreviations: ICD, implantable cardioverter-defibrillator; SD, standard deviation; LVEF, Left Ventricular Ejection Fraction

Table 3
Baseline Characteristics of Patient Groups Discovered via Latent Class Analysis.

Outcome	Group A (N = 82)	Group B (N = 33)	Group C (N = 43)	Group D (N = 76)	*p*-value
	Number of Events (n) per Group				
Medication Changes					
Diuretics Increased	19	2	17	8	0.001
Diuretics Decreased	10	3	13	13	0.018
ARNI Increased	28	6	6	12	0.021
ARNI Decreased	12	0	2	4	0.028
ARNI Prescribed	8	5	1	2	0.037
ARB Decreased	0	0	0	5	0.038
ACE Inhibitor Increased	6	6	1	2	0.048
Biomarker & Lab Changes					
Worsening Renal Function (Creatinine)	31	1	9	12	0.002
High-Sensitivity Troponin T Elevation	47	7	25	39	0.005
Worsening Renal Function (eGFR)	33	2	14	19	0.013
Clinical Events & Patient-Reported Outcomes					
Cardiovascular Hospitalization	28	11	21	9	0.012
Deterioration in KCCQ Score	24	7	21	20	0.017
Heart Failure Hospitalization	41	1	15	18	0.028

Note: Values represent the total number of events (n) observed in each group. P-values are from Kruskal-Wallis tests of normalized outcomes based on patient time (last event − first event) in years. Outcomes are sorted by significance within each category. Abbreviations: KCCQ, Kansas City Cardiomyopathy Questionnaire; ARNI, Angiotensin Receptor-Neprilysin Inhibitor; ARB, Angiotensin II Receptor Blocker; ACE, Angiotensin-Converting Enzyme;

Table 4
Comparison of Baseline Characteristics Between Group A and Group B.

Characteristic	Group A (N = 82)	Group B (N = 33)	*p*-value
Demographics			
Female Sex, n (%)	9 (11.0)	4 (12.1)	>0.99
Age, mean ± SD (years)	72.8 ± 9.9	63.6 ± 11.1	<0.001
LVEF, mean ± SD (%)	33.8 ± 9.4	37.7 ± 8.3	0.04
Cardiovascular History, n (%)			
Myocardial Infarction	57 (69.5)	27 (81.8)	0.25
Percutaneous Coronary Intervention	57 (69.5)	33 (100.0)	<0.001
Coronary Artery Bypass Grafting	37 (45.1)	7 (21.2)	0.02
Peripheral Artery Disease	21 (25.6)	2 (6.1)	0.02
Atrial Fibrillation	32 (39.0)	2 (6.1)	<0.001
Stroke	9 (11.0)	0 (0.0)	0.06
Pacemaker/ICD	47 (57.3)	11 (33.3)	0.02
Other Comorbidities, n (%)			
Hypertension	63 (76.8)	12 (36.4)	<0.001
Diabetes Mellitus	34 (41.5)	0 (0.0)	<0.001
Hypercholesterolemia	22 (26.8)	12 (36.4)	0.50
Chronic Obstructive Pulmonary Disease	24 (29.3)	0 (0.0)	<0.001
Chronic Kidney Disease	53 (64.6)	0 (0.0)	<0.001

Abbreviations: ICD, implantable cardioverter-defibrillator; SD, standard deviation; LVEF, Left Ventricular Ejection Fraction

report and candidate labels. The label with the highest similarity score is selected as the corresponding event.

2.3.4. Ordering of Same-Day events

Because healthcare data frequently contains day-level timestamps, the exact sequence of same-day events is often unavailable. We therefore impose an ordering based on the HF management workflow at the participating clinic. Events are ordered as follows: (1) biomarker and symptom changes, (2) PROs, (3) clinical outcomes, and (4) prescription changes. This ordering reflects how information is generated and documented during care: physiological changes are assessed first, PROs capture the patient's status during the encounter, clinical outcomes are recorded as major events, and prescription records are placed last because they represent the adjusted medication regimen after the encounter or at discharge. Such deterministic tie-breaking is also required in many process-mining settings, since directly-follows relations are otherwise not uniquely defined for events sharing identical timestamps.

2.4. Event abstraction

The enriched event log remains too detailed for effective process discovery, often resulting in highly complex process models [1]. To improve interpretability, we applied a supervised event abstraction approach that maps low-level events to clinically meaningful higher-level concepts, using a knowledge graph. The approach comprises six steps (Fig. 2(a)), of which the first five follow [21].

Domain-specific literature is first collected as the knowledge source. An LLM then generates Competency Questions (CQs), which are reviewed by domain experts and used to construct an ontology by identifying relevant concepts and relationships. The ontology and answers to the validated CQs are subsequently used to build the knowledge graph.

Event abstraction leverages the hierarchical structure of the knowledge graph. Each low-level event is mapped to its corresponding node and replaced by its parent concepts. Events without parents are retained unchanged. This enables multi-level abstraction and process analysis at

Table 5
Comparison of Baseline Characteristics Between Group A and Group D.

Characteristic	Group A (N = 82)	Group D (N = 76)	*p*-values
Demographics			
Female Sex, n (%)	9 (11.0)	27 (35.5)	<0.001
Age, mean ± SD (years)	72.8 ± 9.9	56.8 ± 12.5	<0.001
LVEF, mean ± SD (%)	33.8 ± 9.4	41.2 ± 10.1	<0.001
Cardiovascular History, n (%)			
Myocardial Infarction	57 (69.5)	10 (13.2)	<0.001
Percutaneous Coronary Intervention	57 (69.5)	0 (0.0)	<0.001
Coronary Artery Bypass Grafting	37 (45.1)	6 (7.9)	<0.001
Peripheral Artery Disease	21 (25.6)	0 (0.0)	<0.001
Atrial Fibrillation	32 (39.0)	0 (0.0)	<0.001
Stroke	9 (11.0)	2 (2.6)	0.06
Pacemaker/ICD	47 (57.3)	29 (38.2)	0.02
Other Comorbidities, n (%)			
Hypertension	63 (76.8)	25 (32.9)	<0.001
Diabetes Mellitus	34 (41.5)	10 (13.2)	<0.001
Hypercholesterolemia	22 (26.8)	11 (14.5)	0.05
Chronic Obstructive Pulmonary Disease	24 (29.3)	7 (9.2)	<0.001
Chronic Kidney Disease	53 (64.6)	18 (23.7)	<0.001

Abbreviations: ICD, implantable cardioverter-defibrillator; SD, standard deviation; LVEF, Left Ventricular Ejection Fraction

different levels of granularity, resulting in more interpretable process models (Fig. 2(b)). For more details on event enrichment and abstraction, consult Supplementary Material A.

2.5. Statistical analysis

Baseline demographic and clinical characteristics were evaluated using Fisher's exact test for categorical parameters and using one-way analysis of variance (ANOVA) for continuous variables.

Latent class analysis (LCA) was used to identify discrete, unobserved patient profiles. The optimal number of latent classes was determined using the Akaike information criterion (AIC) [22], and the resulting profiles were reviewed by clinical domain experts to confirm their relevance prior to process model comparison across subgroups.

The identified clusters were further evaluated for differences in clinical activity frequency using the Kruskal–Wallis test as an omnibus test. To account for differences in observation time across patient groups, event frequencies were normalized by patient-time before statistical comparison. To reduce instability in rate estimates caused by very short observation periods, we set a minimum denominator of one year. Where significant differences were detected, Dunn's post-hoc test with Bonferroni correction was applied to identify the specific group pairs driving those differences.

All analyses were implemented in Python using the SciPy library [23], and a two-sided significance threshold of $p < 0.05$ was applied throughout.

3. Results

3.1. General study population

After excluding patients with incomplete follow-up, the final analytical cohort comprised 234 patients with HF. Baseline characteristics are summarized in Table 1. Hypertension was present in 53% of patients and 44% had a pacemaker or an implantable cardioverter-defibrillator (ICD).

3.2. Effectiveness of data enrichment and event abstraction

We evaluated our pipeline against a baseline model constructed from administrative data, where visits are treated as generic outpatient encounters and outcomes are limited to admission or death. As shown in Fig. 3(a) and 3(b), the baseline achieves constantly high fitness across

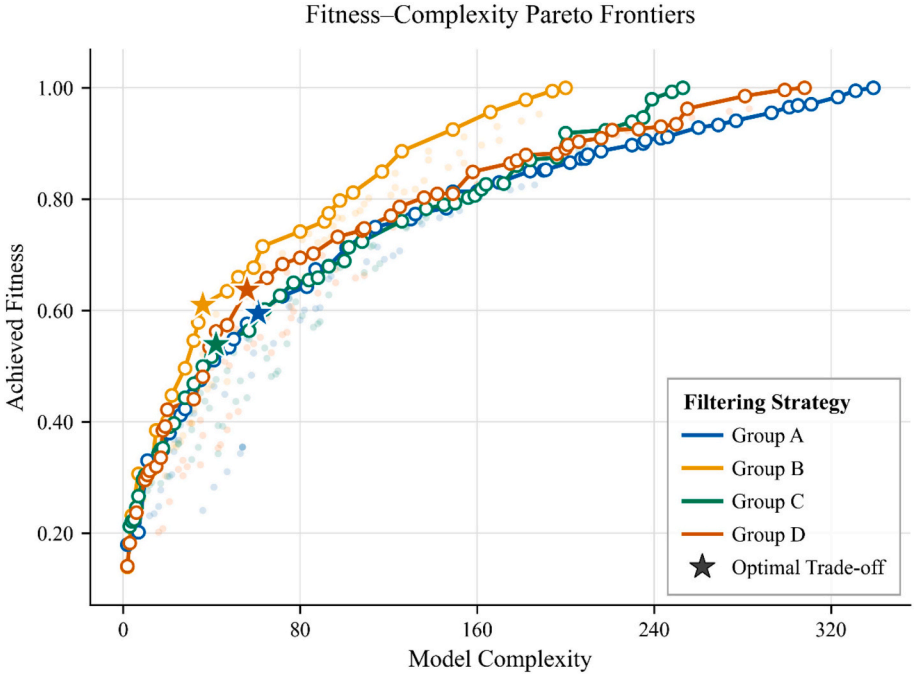

Fig. 4. (a): The fitness-complexity pareto frontier; the further discussed models are marked with star and selected according to the formulation in Section 2.1 Fig. 4 (b): Discovered process model for the patients from Group A after applying one step of event abstraction (fitness: 0.750, raw complexity: 114). Parameters were set to path filtering = 0.7, activity filtering = 0.4. Fig. 4(c): Discovered process model for the patients from Group B after applying one step of event abstraction (fitness: 0.610, raw complexity: 36). Parameters were set to path filtering = 1.0, activity filtering = 0.2. Fig. 4(d): Discovered process model for the patients from Group C after applying one step of event abstraction (fitness: 0.539, raw complexity: 42). Parameters were set to path filtering = 1.0, activity filtering = 0.2. Fig. 4(e): Discovered process model for the patients from Group D after applying one step of event abstraction (fitness: 0.683, raw complexity: 72). Parameters were set to path filtering = 0.9, activity filtering = 0.3.

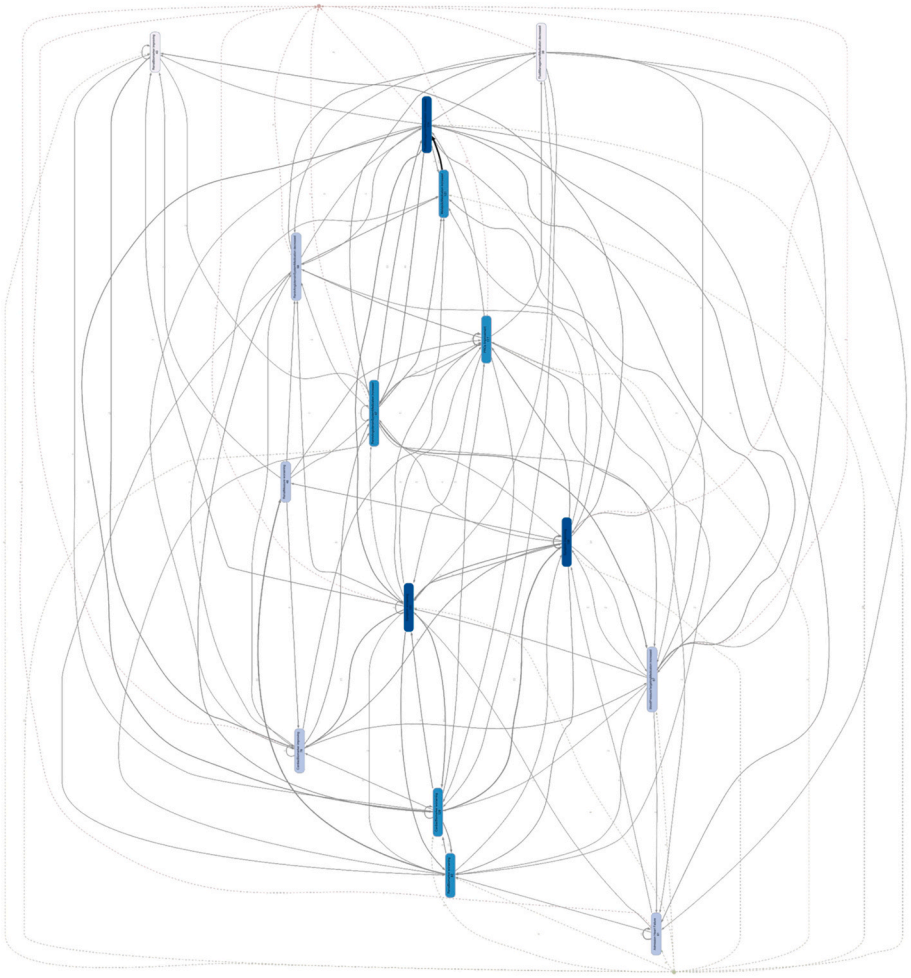

Fig. 4. (*continued*).

filtering thresholds and remains constantly simple, nevertheless, this comes at the cost of limited clinical depth (cf. Fig. 3(c)). In contrast, the enriched log before abstraction contained substantially more clinical detail (Fig. 3(d)), but the models showed an obvious trade-off between complexity and fitness. The abstracted logs with one level of abstraction (L1) and two levels of abstraction (L2) provided a more balanced representation (Fig. 3(e) and (f)), retaining clinically meaningful process information, while improving fitness and simplicity with regard to the enriched log. Overall, abstraction appears to offer a useful compromise between interpretability, conformance, and clinical expressiveness. Additionally, we compared our event abstraction mechanism with a supervised baseline [24]. Compared with the hierarchical-list baseline, KG-based abstraction—particularly at the second abstraction level—maintained high fitness while keeping model complexity stable across filtering thresholds, supporting its use as a robust and semantically grounded abstraction strategy; detailed results are provided in Supplementary Material A.

Further analysis focused on one level of abstraction (L1) because it provided the most useful balance between detail and simplicity. Compared with the original enriched log, L1 reduced complexity and improved model quality, while preserving more clinically meaningful distinctions than L2. Models generated using L2 were simpler, yet their broader abstraction may hide relevant differences.

3.3. Patient Stratification

We identified four patient clusters based on medical history (Table 2). Group A consisted of older, multi-morbid patients (mean age 72.8 years), with high rates of hypertension (76.8%) and chronic kidney disease (64.6%), representing advanced HF. Group B included younger patients with universal prior percutaneous coronary intervention (100%), and frequent prior myocardial infarction (81.8%). Group C comprised older patients (mean age 70.5 years) predominantly with atrial fibrillation (97.7%), indicating arrhythmia-associated HF. Group D represented younger patients with few comorbidities, consistent with early-stage HF.

Comparative analysis of the enriched event log (Table 3) showed significant differences in clinical activity across the clusters. Group A experienced higher rates of heart failure-related hospitalizations compared with Group B (Dunn's test, $p = 0.02$, Table 4) and more frequent worsening renal function compared with Groups B and D (Dunn's test, $p < 0.001$, Table 4; Dunn's test, $p = 0.03$, Table 5, respectively), consistent with its higher burden of chronic kidney disease.

Process models further illustrated differences in care pathways (Fig. 4(a)-(e)). The process models were selected following the principle discussed in Section 2.1.

Group A showed the most interconnected structure, with frequent transitions among medication adjustments, symptom changes, biomarker changes, and PRO-related events. This was consistent with its high comorbidity burden and the higher frequency of heart-failure hospitalization and renal-function deterioration observed in the activity-frequency analysis.

Groups B and C showed more focused pathway structures. In Group B, transitions were mainly concentrated around medication management, symptoms, PROs, and cardiac biomarker worsening, consistent

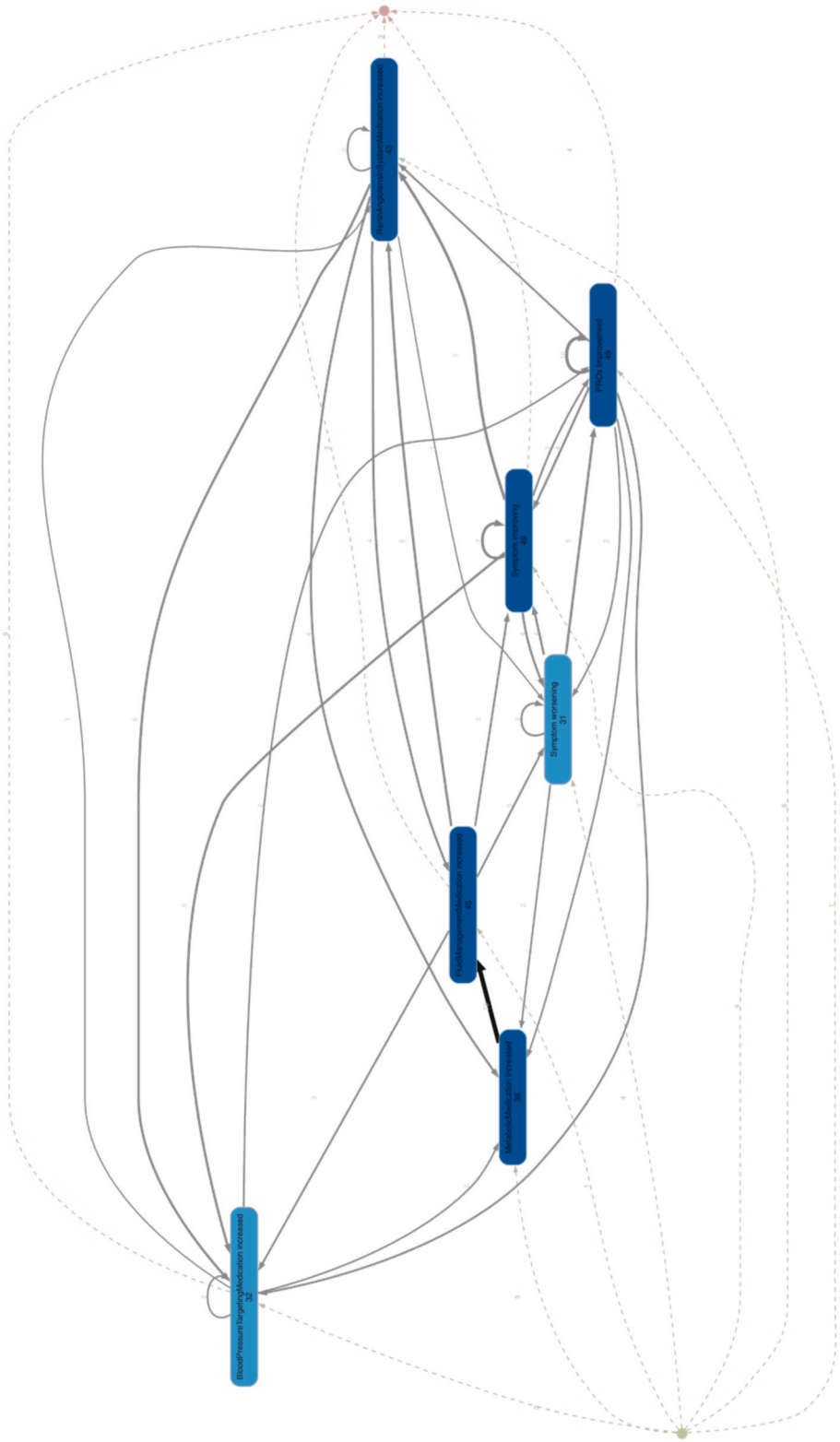

Fig. 4. (*continued*).

with its ischemic cardiovascular profile. Group C was similarly structured around symptoms, PROs, and cardiac biomarker events, aligning with its arrhythmia-associated phenotype.

Group D, although younger and less comorbid, still showed recurrent transitions involving symptoms, PROs, biomarkers, and medication-management events. Thus, lower baseline comorbidity did not necessarily correspond to a simple longitudinal care pathway.

Overall, the subgroup models suggested that enriched and abstracted event logs captured phenotype-specific differences in pathway organization: Group A was most diffuse, Groups B and C were more focused, and Group D remained clinically heterogeneous despite lower baseline burden. For additional patient stratification information and results, we refer the reader to Supplementary Material B.

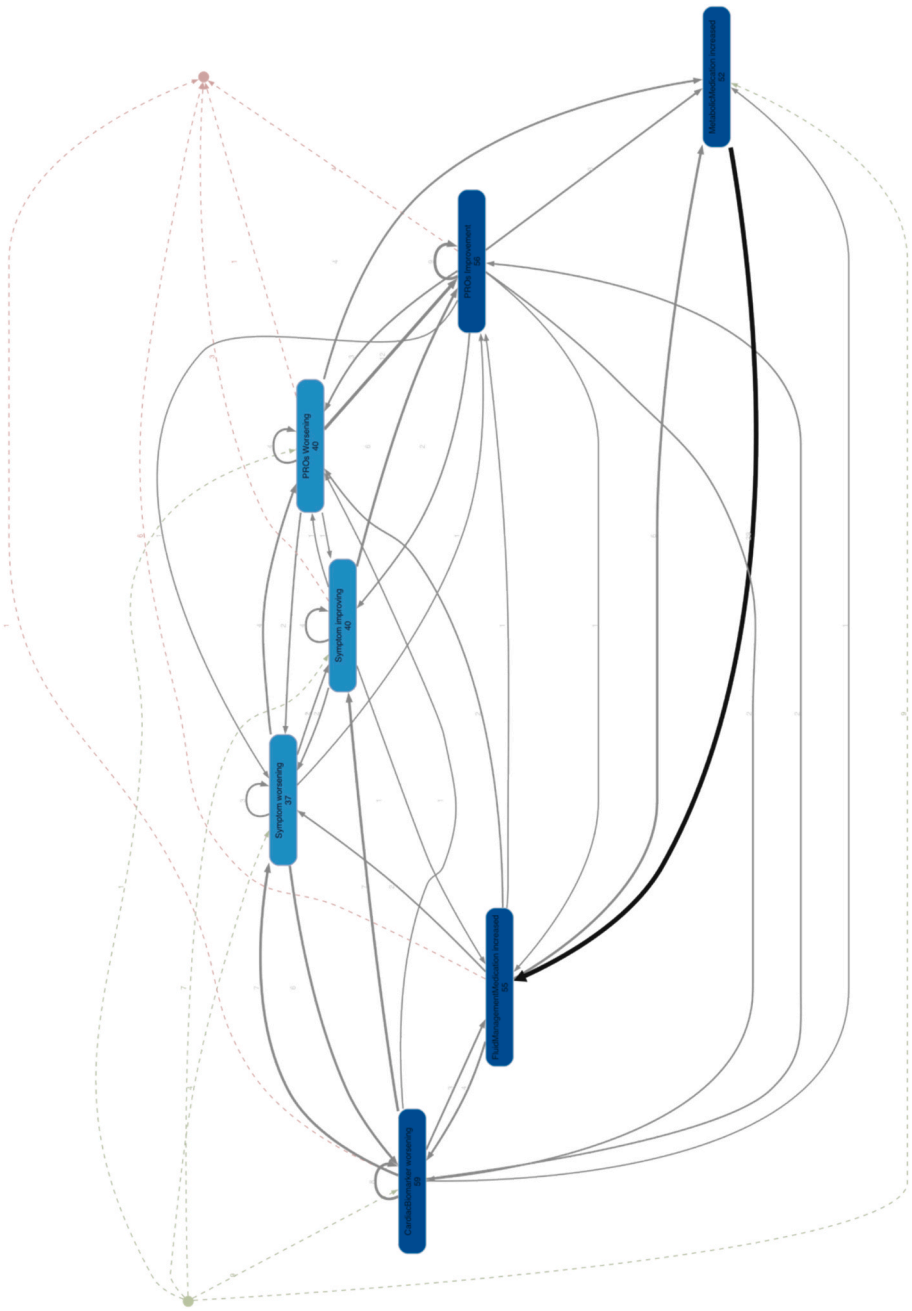

Fig. 4. (*continued*).

4. Discussion

This paper introduces and validates a novel two-stage pipeline to construct interpretable, patient-centric process models from multidimensional treatment data. Our primary finding is that the proposed approach reveals clinical insights from data often overlooked in conventional process mining. Reflecting the heterogeneity in age, comorbidities, and disease progression, we identified four distinct clusters. Our approach captured key differences in treatment pathways among these cohorts, including variations in hospitalization rates, renal function deterioration, and medication management patterns.

The evaluation results demonstrate that the method is a significant step beyond previous process mining attempts to analyze treatment pathways. Findings reveal that older, multi-morbid patients in Group A experienced higher rates of HF–related hospitalizations and worsening renal function compared to other groups. Process models for this cohort exhibited complex, interconnected structures with multiple medication adjustment loops — characteristic of fragile patients requiring continuous therapeutic management. In contrast, younger or less comorbid patients (Groups B, C, and D) followed more linear and predictable pathways, reflecting more stable disease trajectories. Recurrent medication adjustment loops were particularly evident in ischemic patients, highlighting a reactive, trial-and-error approach to therapy. These observations demonstrate how our pipeline not only identifies at-risk patients but also elucidates how risks manifest within treatment pathways.

Compared with prior work on the same dataset [11], our method enables a deeper, pathway-level analysis by integrating previously excluded clinical dynamics from textual reports and patient-reported measures. Unlike oncology-focused approaches [8] that rely on coarse abstractions, our knowledge-graph-based methodology preserves clinically relevant detail, allowing identification of specific patterns like iterative medication loops. Similarly, while theoretical frameworks have

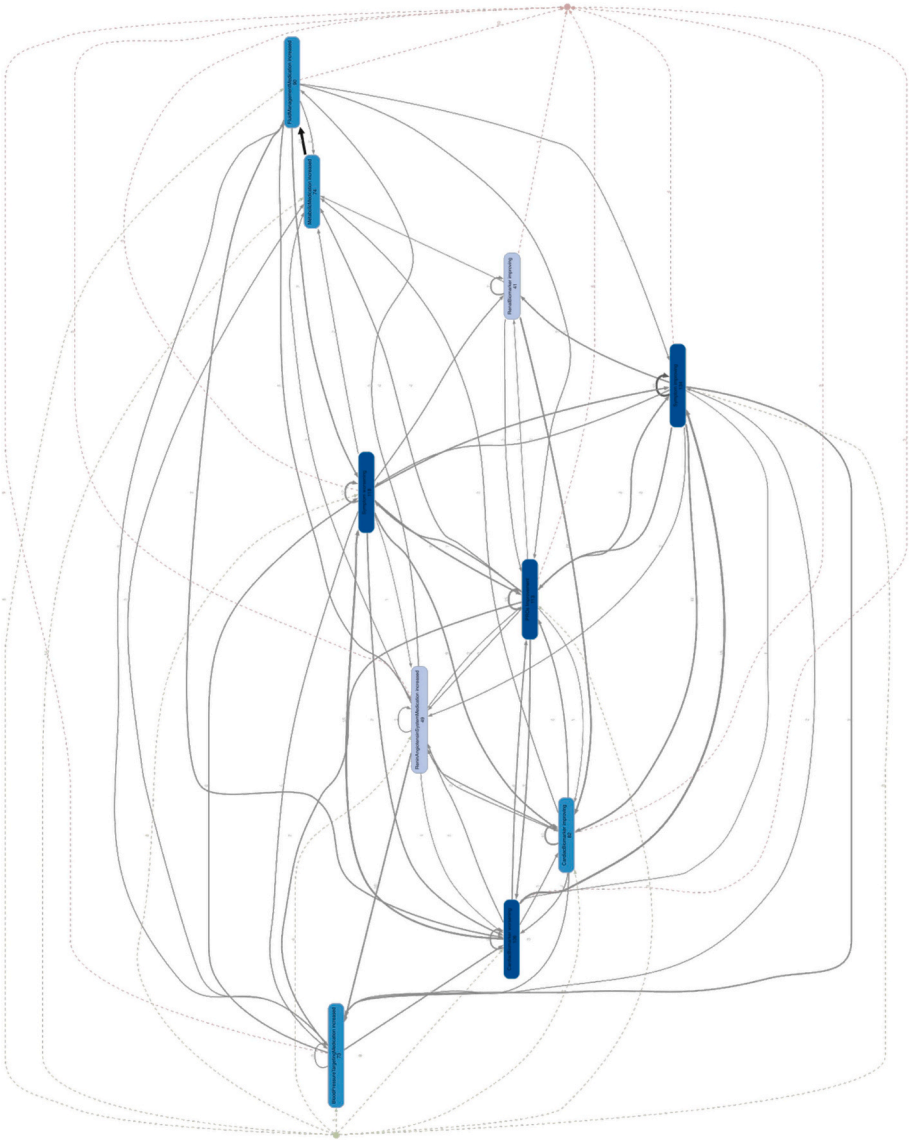

Fig. 4. (*continued*).

proposed integrating patient-centered measures into process discovery [25,26], our pipeline provides a practical, fully implementable methodology within the process mining paradigm, without necessitating a shift from the case-centric perspective.

The primary strength of the presented approach is its end-to-end pipeline that bridges raw, multi-modal clinical data and interpretable process models, with its applicability demonstrated on real-world clinical data. However, we acknowledge several limitations. Our method assumes complete data, and future work should explore robust imputation strategies. Additionally, the knowledge graph for LLM-supported event abstraction was human-reviewed for clinical plausibility and source-data consistency, nevertheless, formal quantitative validation of LLM performance was outside the scope of this study. Furthermore, the small sample size (N = 234) and single-center design mean our findings should be considered a methodological demonstration, as the discovered patient phenotypes may be sample-specific and require further validation in larger, multi-center cohorts to ensure generalizability. Overall, the proposed pipeline represents a significant advance in patient-centered process mining, enabling both quantitative and qualitative insights into treatment pathways that were previously inaccessible.

5. Conclusion

In conclusion, this study demonstrates that the generation of clinically meaningful events reflecting substantive changes in patient treatment is essential for the effective application of process mining in healthcare settings, in particular heart failure. Validation across four distinct heart failure cohorts revealed significant inter-cohort differences in hospitalization rates, renal function deterioration, and medication management patterns that conventional process mining analyses fail to surface. These findings, corroborated by clinical domain experts, demonstrate that the proposed pipeline offers a practical and reproducible framework for translating complex, multi-modal clinical data into interpretable and actionable insights. Overall, this work represents a step toward more data-driven, patient-centered healthcare and highlights the potential of process mining to inform personalized treatment strategies.

CRediT authorship contribution statement

Anton Antonov: Writing – review & editing, Writing – original draft, Visualization, Software, Investigation, Formal analysis, Conceptualization. **Harry H. Beyel:** Writing – review & editing, Writing – original

draft, Supervision, Data curation, Conceptualization. **Marlo Verket:** Writing – review & editing, Writing – original draft, Supervision, Data curation, Conceptualization. **Christopher T. Schwanen:** Writing – review & editing. **Viki Peeva:** Writing – review & editing. **Marco Pegoraro:** Writing – review & editing. **Julia Brandts:** Writing – review & editing. **Dirk Müller-Wieland:** Writing – review & editing. **Nikolaus Marx:** Writing – review & editing. **Wil M.P. van der Aalst:** Writing – review & editing. **Katharina Marx-Schütt:** Writing – review & editing.

Funding

This study was funded by the Department of Internal Medicine I, RWTH Aachen University Hospital.

Data Availability

The authors do not have permission to share data due to regulations on sensitive patient information.

Code Availability

Code access may be granted upon reasonable written request and subject to institutional approval.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

N. Marx is supported by the Deutsche Forschungsgemeinschaft (German Research Foundation; TRR 219; Project-ID 322900939 [M03, M05]). K. Marx-Schütt is supported by the Deutsche Forschungsgemeinschaft (German Research Foundation; TRR 219; Project-ID 322900939 [C07]).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijmedinf.2026.106601.

References

[1] W.M.P. Van Der Aalst. Process Mining - Data Science in Action, Springer Berlin Heidelberg, Berlin, Heidelberg, 2016, https://doi.org/10.1007/978-3-662-49851-4.
[2] E. De Roock, N. Martin, Process mining in healthcare – an updated perspective on the state of the art, J. Biomed. Inform. 127 (Mar. 2022) 103995, https://doi.org/10.1016/j.jbi.2022.103995.
[3] E. Rojas, J. Munoz-Gama, M. Sepúlveda, D. Capurro, Process mining in healthcare: a literature review, J. Biomed. Inform. 61 (Jun. 2016) 224–236, https://doi.org/10.1016/j.jbi.2016.04.007.
[4] M. Packer, M. Metra, Guideline-Directed Medical Therapy for Heart failure does not Exist: a Non-Judgmental Framework for describing the Level of Adherence to Evidence-based Drug Treatments for patients with a Reduced Ejection Fraction, Eur. J. Heart Fail. 22 (10) (Oct. 2020) 1759–1767, https://doi.org/10.1002/ejhf.1857.
[5] C.J. Hayward, et al., Disease trajectories following myocardial infarction: insights from process mining of 145 million hospitalisation episodes, eBioMedicine 96 (Oct. 2023) 104792, https://doi.org/10.1016/j.ebiom.2023.104792.
[6] N. Martin, A. Martinez-Millana, B. Valdivieso, and C. Fernández-Llatas, Interactive Data Cleaning for Process Mining: A Case Study of an Outpatient Clinic's Appointment System, in Business Process Management Workshops, vol. 362, C. Di Francescomarino, R. Dijkman, and U. Zdun, Eds., in Lecture Notes in Business Information Processing, vol. 362, Cham: Springer International Publishing, 2019, pp. 532–544. doi: 10.1007/978-3-030-37453-2_43.
[7] E. Benevento et al., Process Modeling and Conformance Checking in Healthcare: A COVID-19 Case Study, in Process Mining Workshops, vol. 468, M. Montali, A. Senderovich, and M. Weidlich, Eds., in Lecture Notes in Business Information Processing, vol. 468, Cham: Springer Nature Switzerland, 2023, pp. 315–327. doi: 10.1007/978-3-031-27815-0_23.
[8] R. P. J. C. Bose and W. M. P. Van Der Aalst, Analysis of Patient Treatment Procedures, in Business Process Management Workshops, vol. 99, F. Daniel, K. Barkaoui, and S. Dustdar, Eds., in Lecture Notes in Business Information Processing, vol. 99. , Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 165–166. doi: 10.1007/978-3-642-28108-2_17.
[9] M. Arias, et al., Mapping the Patient's Journey in Healthcare through Process Mining, Int. J. Environ. Res. Public Health 17 (18) (Sep. 2020) 6586, https://doi.org/10.3390/ijerph17186586.
[10] J. Cremerius, H. Patzlaff, and M. Weske, Relationships Between Change Patterns in Dynamic Event Attributes, in *Business Process Management Workshops*, vol. 492, J. De Weerdt and L. Pufahl, Eds., in Lecture Notes in Business Information Processing, vol. 492. , Cham: Springer Nature Switzerland, 2024, pp. 149–160. doi: 10.1007/978-3-031-50974-2_12.
[11] H. Beyel, et al., in: Process-Aware Analysis of Treatment Paths in Heart Failure Patients: A Case Study, SCITEPRESS - Science and Technology Publications, Rome, Italy, 2024, pp. 506–515, https://doi.org/10.5220/0012392600003657.
[12] A. Berti, S. Van Zelst, D. Schuster, PM4Py: a process mining library for Python, Softw. Impacts 17 (Sep. 2023) 100556, https://doi.org/10.1016/j.simpa.2023.100556.
[13] E.W. Myers, W. Miller, Approximate matching of regular expressions, Bull. Math. Biol. 51 (1) (Jan. 1989) 5–37, https://doi.org/10.1007/BF02458834.
[14] D. Schuster, S. Van Zelst, and W. M. P. Van Der Aalst, Alignment Approximation for Process Trees, in *Process Mining Workshops*, vol. 406, S. Leemans and H. Leopold, Eds., in Lecture Notes in Business Information Processing, vol. 406. , Cham: Springer International Publishing, 2021, pp. 247–259. doi: 10.1007/978-3-030-72693-5_19.
[15] X. Lu, A. Gal, H.A. Reijers. Discovering Hierarchical Processes Using Flexible Activity Trees for Event Abstraction, IEEE, Padua, Italy, Oct. 2020, pp. 145–152, https://doi.org/10.1109/ICPM49681.2020.00030.
[16] T.A. McDonagh, et al., 2021 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure, Eur. Heart J. 42 (36) (Sep. 2021) 3599–3726, https://doi.org/10.1093/eurheartj/ehab368.
[17] C.P. Green, C.B. Porter, D.R. Bresnahan, J.A. Spertus, Development and evaluation of the Kansas City Cardiomyopathy Questionnaire: a new health status measure for heart failure, J. Am. Coll. Cardiol. 35 (5) (Apr. 2000) 1245–1255, https://doi.org/10.1016/S0735-1097(00)00531-3.
[18] A.S. Zigmond, R.P. Snaith, The Hospital anxiety and Depression Scale, Acta Psychiatr. Scand. 67 (6) (Jun. 1983) 361–370, https://doi.org/10.1111/j.1600-0447.1983.tb09716.x.
[19] R.Y. Maragheh, et al.. LLM-TAKE: Theme-Aware Keyword Extraction Using Large Language Models, IEEE, Sorrento, Italy, Dec. 2023, pp. 4318–4324, https://doi.org/10.1109/BigData59044.2023.10386476.
[20] F. Remy, K. Demuynck, T. Demeester, BioLORD: Learning Ontological Representations from Definitions for Biomedical Concepts and their Textual Descriptions, in Findings of the Association for Computational Linguistics: EMNLP, Abu Dhabi, United Arab Emirates: Association for, Comput. Linguist. 2022 (2022) 1454–1465, https://doi.org/10.18653/v1/2022.findings-emnlp.104.
[21] V. K. Kommineni, B. König-Ries, and S. Samuel, From human experts to machines: An LLM supported approach to ontology and knowledge graph construction, 2024, *arXiv*. doi: 10.48550/ARXIV.2403.08345.
[22] H. Akaike, A new look at the statistical model identification, IEEE Trans. Autom. Control 19 (6) (Dec. 1974) 716–723, https://doi.org/10.1109/TAC.1974.1100705.
[23] P. Virtanen, et al., SciPy 1.0: fundamental algorithms for scientific computing in Python, Nat. Methods 17 (3) (Mar. 2020) 261–272, https://doi.org/10.1038/s41592-019-0686-2.
[24] F. Klessascheck, et al., Domain-Specific Event Abstraction, *Bus*, Inf. Syst. (Jul. 2021) 117–126, https://doi.org/10.52825/bis.v1i.39.
[25] M. Naeimaei Aali, F. Mannhardt, and P. Jelle Toussaint, Clinical Event Knowledge Graphs: Enriching Healthcare Event Data with Entities and Clinical Concepts - Research Paper, in *Process Mining Workshops*, vol. 503, J. De Smedt and P. Soffer, Eds., in Lecture Notes in Business Information Processing, vol. 503. , Cham: Springer Nature Switzerland, 2024, pp. 296–308. doi: 10.1007/978-3-031-56107-8_23.
[26] M. Naeimaei Aali, F. Mannhardt, and P. J. Toussaint, Discovering Care Pathways for Multi-morbid Patients Using Event Graphs, in *Process Mining Workshops*, vol. 433, J. Munoz-Gama and X. Lu, Eds., in Lecture Notes in Business Information Processing, vol. 433. , Cham: Springer International Publishing, 2022, pp. 352–364. doi: 10.1007/978-3-030-98581-3_26.