Prozessbasierte Analyse und Interpretation onkologischer Behandlungsverläufe mit Machine Learning

Digitalisierung im Gesundheitswesen – Neue Perspektiven

Flavio Horbach M.Sc. | Jana Vormann M.Sc.

Prof. Dr. Sven Pagel | Prof. Dr. Tobias Walter

universität koblenz
Informatik

Rheinland-Pfalz
MINISTERIUM FÜR
WISSENSCHAFT
UND GESUNDHEIT

HOCHSCHULE MAINZ
UNIVERSITY OF
APPLIED SCIENCES

23/04/2025 13:05 – 13:35

Agenda

1. Vorstellung Forschungskolleg AI-DPA

2. Kurzeinführung Process Mining

Insights on Prostate Cancer Treatment Pathways Using Process Discovery

4. Predictive Process Monitoring of Treatment Pathways

5. Ausblick

universität koblenz
Informatik

RheinlandPfalz
MINISTERIUM FÜR WISSENSCHAFT UND GESUNDHEIT

HOCHSCHULE MAINZ
UNIVERSITY OF APPLIED SCIENCES

1. Vorstellung Forschungskolleg AI-DPA

23/04/2025 13:05 – 13:35

Unser Forschungskolleg

Was ist AI-DPA?

Analyse und **I**nterpretation von unstrukturierten **D**aten und **P**rozessen in zwei- und dreidimensionalen **A**nwendungsszenarien mit Machine Learning

v. links: Prof. Dr. Thomas Klauer, Univ.-Prof. Dr. Patrick Delfman, Prof. Dr. Tobias Walter, Janine Buchholz, Jonas Blatt, Jessica Buchner, Prof. Dr. Sven Pagel, Jana Vormann, Prof. Dr. Peer Neubert, Flavio Horbach (es fehlt Univ.-Prof. Dr. Dietrich Paulus)

Ziele und Datenpartner

- Prozessanalyse über verschiedene Domänen (u.a. Medien, Gesundheitswesen) hinweg
- Entwicklung von KI-Algorithmen
- Verbesserung der Datenzugänglichkeit
- Prozessvorhersage und -verbesserung

Block 1: **Semantische Interpretation von unstrukturierten, kolorierten 3D-Punktwolken in barrierefreien Gebäuden/umbaute Bereichen (Indoor/Outdoor)** Univ.-Prof. Dr. Dietrich Paulus / Univ.-Prof. Dr. Peer Neubert, Prof. Dr. Thomas Klauer		
Promotion 1-1 „Robuste Registrierung und semantische Interpretation von Punktwolken im Außenbereich" (**Paulus/Neubert**)	Promotion 1-2 „Adaptive Registrierung und semantische Interpretation von Punktwolken in Innenräumen" (**Klauer**)	
Block 2: **Predictive Process Monitoring für unstrukturierte Prozesse in barrierefreien Medien** Univ.-Prof. Dr. Patrick Delfmann, Prof. Dr. Sven Pagel, Prof. Dr. Tobias Walter		
Promotion 2-1 „Predictive Process Monitoring in Chatbots zur Analyse von Antwortenpfaden" (technologieorientiert) (**Delfmann**)	Promotion 2-2 „Predictive Process Monitoring von Click-Daten zur Analyse und Vorhersage des Nutzerverhalten" (technologieorientiert) (**Walter**)	Promotion 2-3 „Predictive Process Monitoring von semi- und unstrukturierten Prozessen in Medienunternehmen" (wirtschaftsorientiert) (**Pagel**)

Unsere Datenpartner

ARD 1

Klickdaten zu Concent-Nutzung, Impressions, Teaser-Klicks

IDG INSTITUT FÜR DIGITALE GESUNDHEITSDATEN

Krebsregister Rheinland-Pfalz im Institut für digitale Gesundheitsdaten

Behandlungsverläufe zu Krebserkrankungen

Mehr unter: ai-dpa.hs-mainz.de

Ziele und Datenpartner

- Prozessanalyse über verschiedene Domänen (u.a. Medien, Gesundheitswesen) hinweg
- Entwicklung von KI-Algorithmen
- Verbesserung der Datenzugänglichkeit
- Prozessvorhersage und -verbesserung

Block 1: **Semantische Interpretation von unstrukturierten, kolorierten 3D-Punktwolken in barrierefreien Gebäuden/umbaute Bereichen (Indoor/Outdoor)** Univ.-Prof. Dr. Dietrich Paulus / Univ.-Prof. Dr. Peer Neubert, Prof. Dr. Thomas Klauer		
Promotion 1-1 „Robuste Registrierung und semantische Interpretation von Punktwolken im Außenbereich“ (**Paulus/Neubert**)	Promotion 1-2 „Adaptive Registrierung und semantische Interpretation von Punktwolken in Innenräumen“ (**Klauer**)	
Block 2: **Predictive Process Monitoring für unstrukturierte Prozesse in barrierefreien Medien** Univ.-Prof. Dr. Patrick Delfmann, Prof. Dr. Sven Pagel, Prof. Dr. Tobias Walter		
Promotion 2-1 „Predictive Process Monitoring in Chatbots zur Analyse von Antwortenpfaden“ (technologieorientiert) (**Delfmann**)	Promotion 2-2 „Predictive Process Monitoring von Click-Daten zur Analyse und Vorhersage des Nutzerverhalten“ (technologieorientiert) (**Walter**)	Promotion 2-3 „Predictive Process Monitoring von semi- und unstrukturierten Prozessen in Medienunternehmen“ (wirtschaftsorientiert) (**Pagel**)

Unsere Datenpartner

Klickdaten zu Concent-Nutzung, Impressions, Teaser-Klicks

Behandlungsverläufe zu Krebserkrankungen

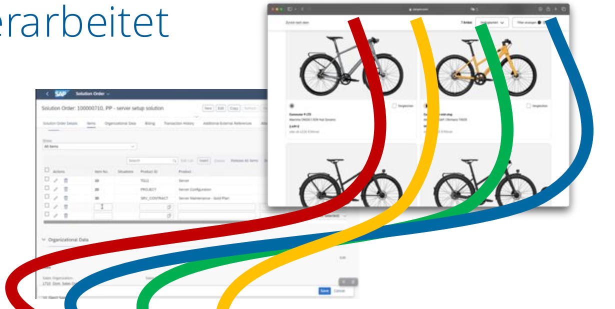
Mehr unter: ai-dpa.hs-mainz.de

2. Kurzeinführung Process Mining

23/04/2025 13:05 – 13:35

Abläufe und Prozesse erkennen

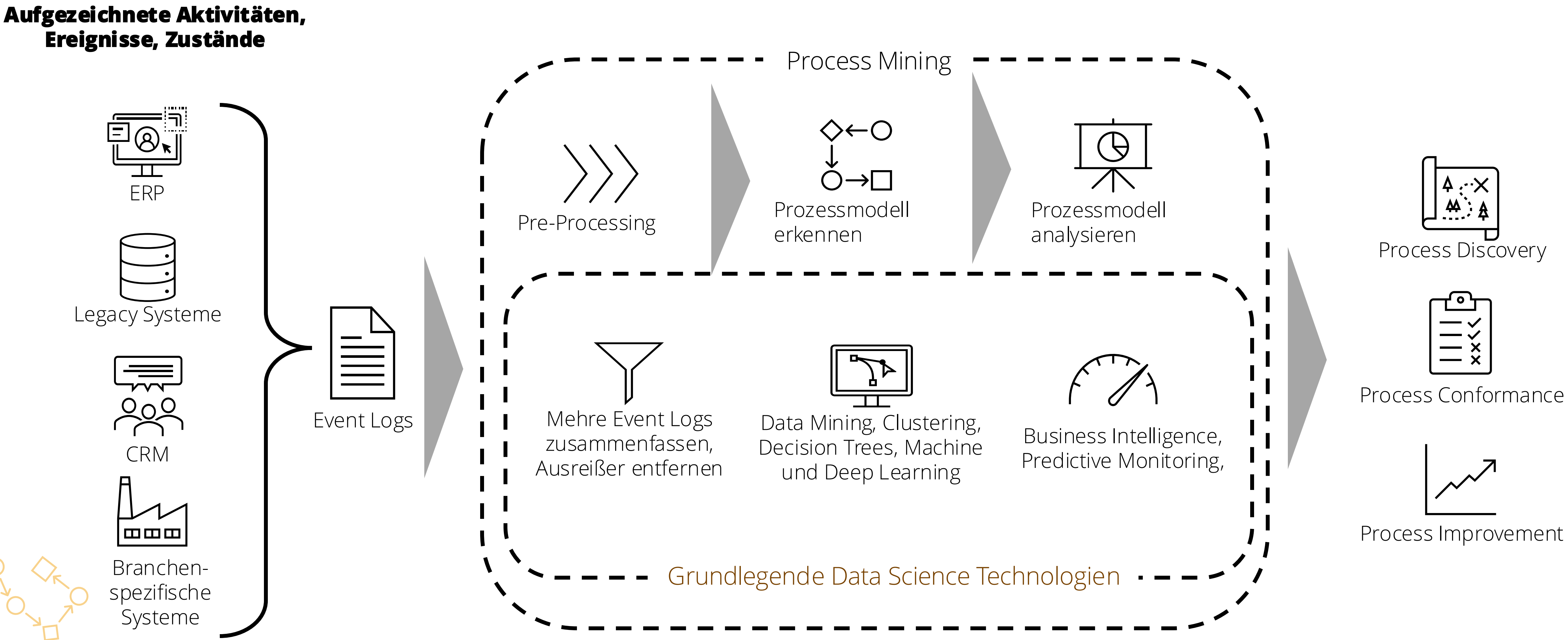
Bestellung wird erstellt und verarbeitet



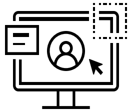
Einzelteile werden beschafft und zum Fahrrad zusammengesetzen

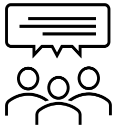
Fertiges Fahrrad wird versendet

Process Mining Technologie



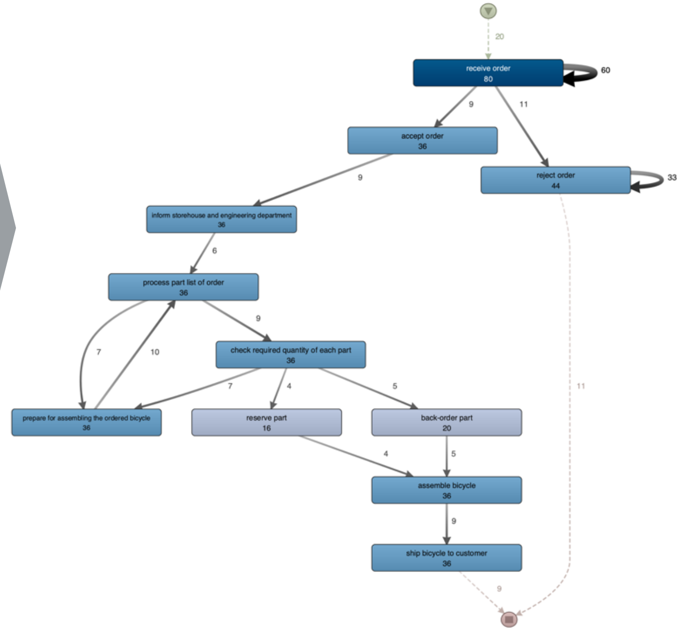
Beispiel eines Event Logs


ERP

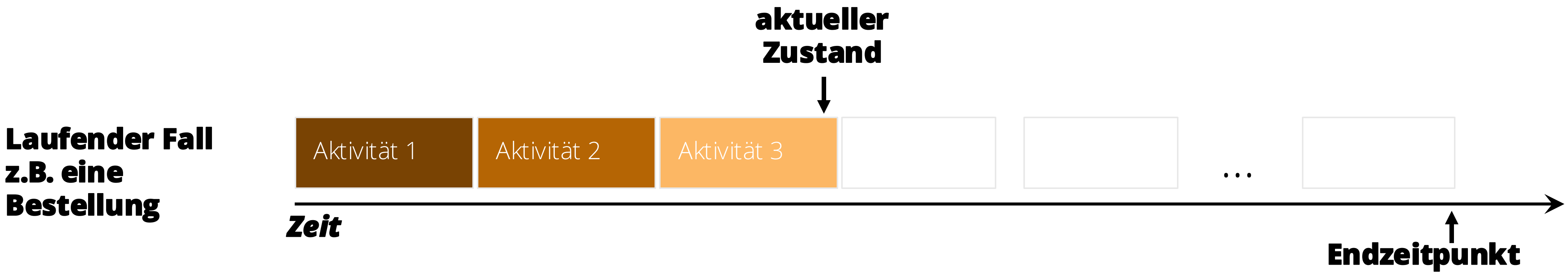

CRM


Event Log

case id	activity name	timestamp
…	…	…
b02794fe	receive order	2023-03-20T14:55:26.803+01:00
b02794fe	accept order	2023-03-21T17:46:27.862+01:00
b02794fe	inform storehouse and engineering department	2023-03-21T17:46:29.881+01:00
b02794fe	prepare for assembling the ordered bicycle	2023-03-22T11:37:33.901+01:00
b02794fe	process part list of order	2023-03-22T17:46:33.899+01:00
b02794fe	check required quantity of each part	2023-03-23T17:50:36.922+01:00
b02794fe	process part list of order	2023-03-23T17:46:36.915+01:00
b02794fe	prepare for assembling the ordered bicycle	2023-03-25T17:46:38.914+01:00
b02794fe	check required quantity of each part	2023-03-25T17:46:43.932+01:00
b02794fe	back-order part	2023-03-26T17:46:43.938+01:00
b02794fe	assemble bicycle	2023-03-27T17:46:52.949+01:00
b02794fe	back-order part	2023-03-27T17:44:52.944+01:00
b02794fe	assemble bicycle	2023-03-28T17:58:59.958+01:00
b02794fe	ship bicycle to customer	2023-03-29T17:46:59.964+01:00
…	…	…

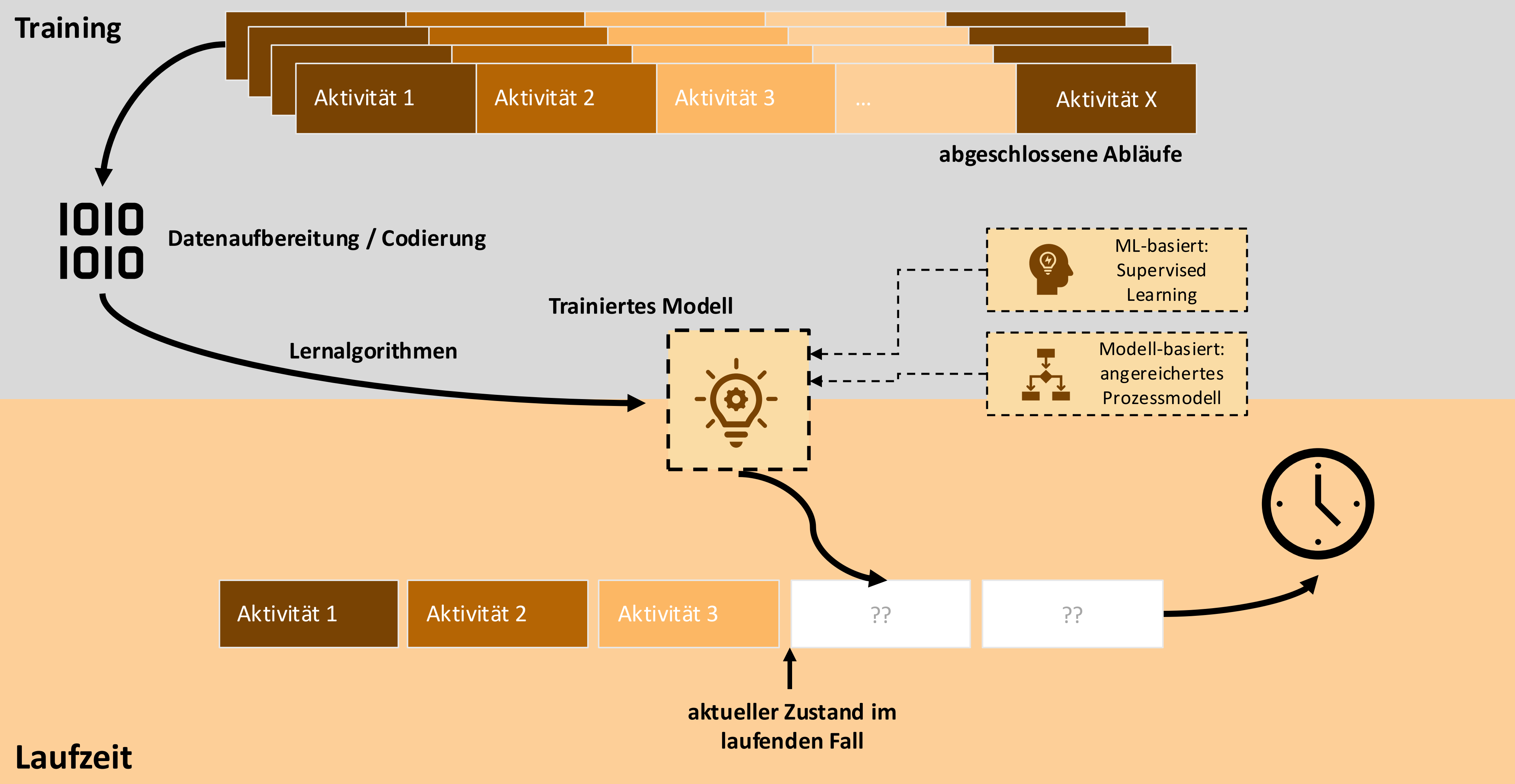


Typische Fragestellungen in laufenden Fällen



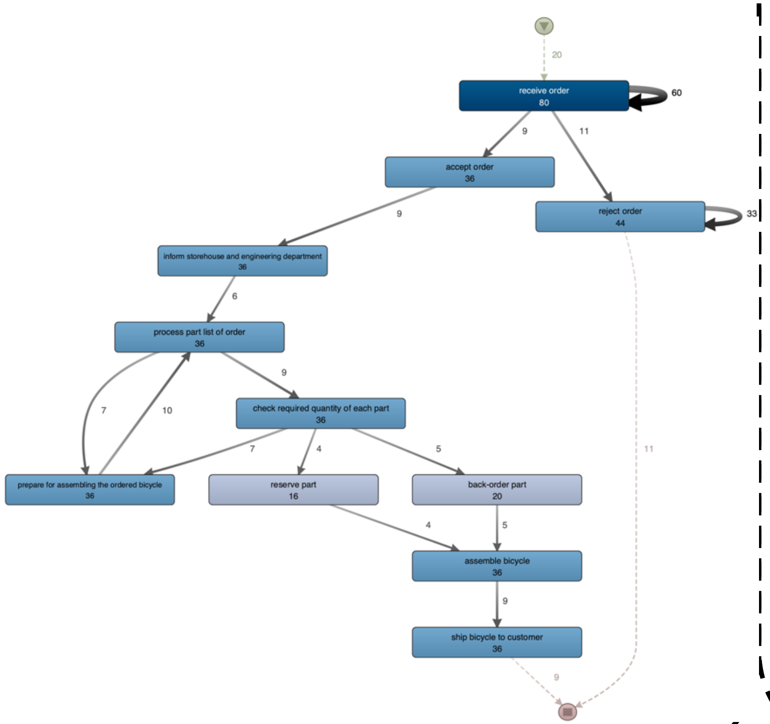
- Was ist die nächste Aktivität des Ablaufs?
  - z.B. wird die Bestellung angenommen oder abgelehnt?
- Wann wird die nächste Aktivität ausgeführt?
  - z.B. wann wird die Rechnung beglichen
- Wie lange läuft der Ablauf bis er zu Ende ist?
  - z.B. wie lange dauert es bis die Ware beim Kunden ankommt?
- Was ist das Ergebnis des Ablaufs?
  - z.B. endet die Bestellung mit einer Reklamation?

Predictive Process Monitoring

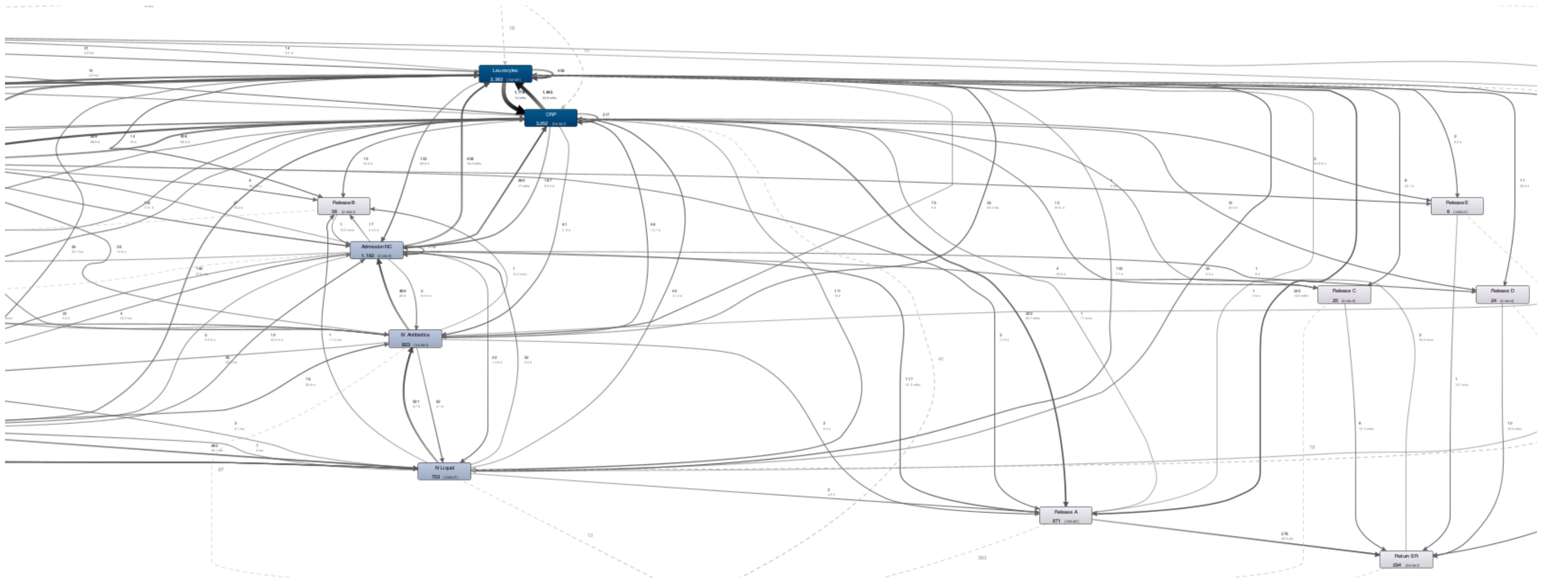


Neue Herausforderungen

Der ideale Ablauf:



Der reale Ablauf,
hier am Beispiel mit 1000 Fällen und 15000 Ereignissen:



Wie behält man hier den Überblick? Wie kann man genaue Vorhersagen machen? → die Aufgabe von AI-DPA

universität koblenz
Informatik

Rheinland-Pfalz
MINISTERIUM FÜR WISSENSCHAFT UND GESUNDHEIT

HOCHSCHULE MAINZ
UNIVERSITY OF APPLIED SCIENCES

Insights on Prostate Cancer Treatment Pathways Using Process Discovery

23/04/2025 13:05 – 13:35

a) ProcessMining4Healthcare:
   a) Oncology
   b) Prostate Carcinoma
b) Method: PM^2
c) Data Collection
d) Data Processing
e) Process Discovery
f) Process Improvement and Support

a) ProcessMining4Healthcare: Oncology

35 million new cases per year (by 2050)[1]

2nd leading cause of death globally[1]

Primary research field in PM4H domain[2]

- Previous research in the **process mining** domain: mainly data from a single institution [3]
- This **case study**:
  - New real-life data set of prostate cancer patients in Rhineland-Palatinate, Germany
  - Health records of multiple institutions (ambulatory and clinical care, pathology)
  - **Aim:** discover and visualize end-to-end patient's pathways and treatments
  - Presented at International Conference on Process Mining 2024 Copenhagen, Denmark
  - Extended journal version for Process Science Journal in review

[1] https://www.who.int/news/item/01-02-2024-global-cancer-burden-growing–amidst-mounting-need-for-services
[2] Chen et al., 2023
[3] Dallagassa et al., 2022; Martin et al., 2022; Munoz-Gama et al., 2022

a) ProcessMining4Healthcare: Prostate Carcinoma

Highest incident rate

Highest prevalence rate

2nd highest mortality rate

… of all cancerous diseases among men in Rhineland-Palatinate and Germany. [1]

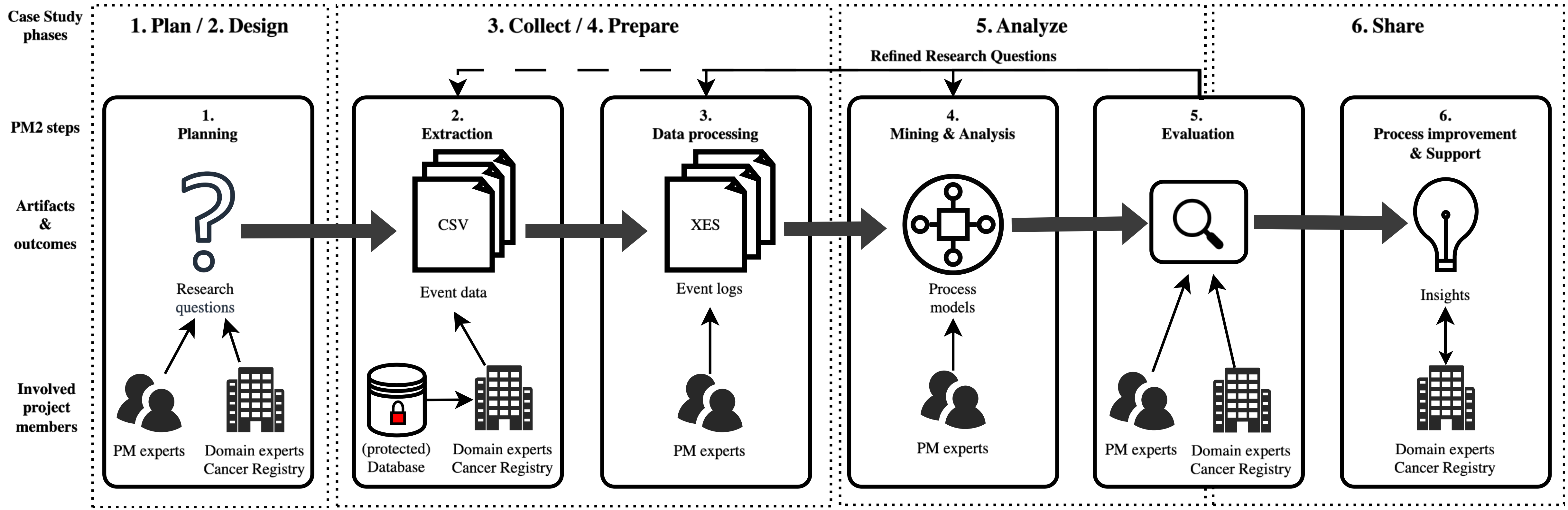
- 5-year survival rate > 90% [2]
- Ambiguous nature of treatment guidelines

→ Large amount of data on treatment of patients

→ Long, eventually complex & varied treatment pathways

[1] https://onkologiemonitor.de, Ronckers et al., 2023
[2] Robert Koch Institut, 2023

b) Method: PM² [1]



- Researchers integrated into IDG organization
- Data processed and analyzed exclusively in IDGs internal network

- Access via virtual machine (Python environment)
- Data processing and analysis with PM4Py

[1] van Eck et al., 2015, Yin, 2014

c) Data Collection – Cancer Registries

- Germany's ***Federal Cancer Early Detection and Cancer Registries Act***[1] & ***State Cancer Registry Act***[2]
  - Obligation of medical institutions to submit data on a patients' cancer treatment to a cancer registry
- Collected data:
  - Diagnosis, treatments, disease status
  - Ambulatory care, clinical care, pathology
- Submission:
  - Online - manually / automatically via standardized interfaces
  - Standardized data format "oncological base data set" [3]
- Cancer Registry Rhineland-Palatinate:
  - Part of Institute for Digital Health Data (IDG), Rhineland-Palatinate [4]
  - Data on prostate cancer patients treated in or living in RPL

196,187
records

2016 – 2025*

26,766
patients

[1] https://dip.bundestag.de/vorgang/gesetz-zur-weiterentwicklung-der-krebsfrüherkennung-und-zur-qualitätssicherung-durch-klinische/47081
[2] https://landesrecht.rlp.de/bsrp/document/jlr-KrebsRegGRP2015rahmen
[3] https://www.basisdatensatz.de
[4] https://www.krebsregister-rlp.de/

c) Data Collection - Partial Data Sets

Systematic Therapy
21,423

Patient
26,766

Diagnosis
29,794

Records
196,187

Radiation
14,929

Tumor Conference
13,990

Follow-up
82,155

Surgery
19,240

Histology
61,797

Pathology
22,127

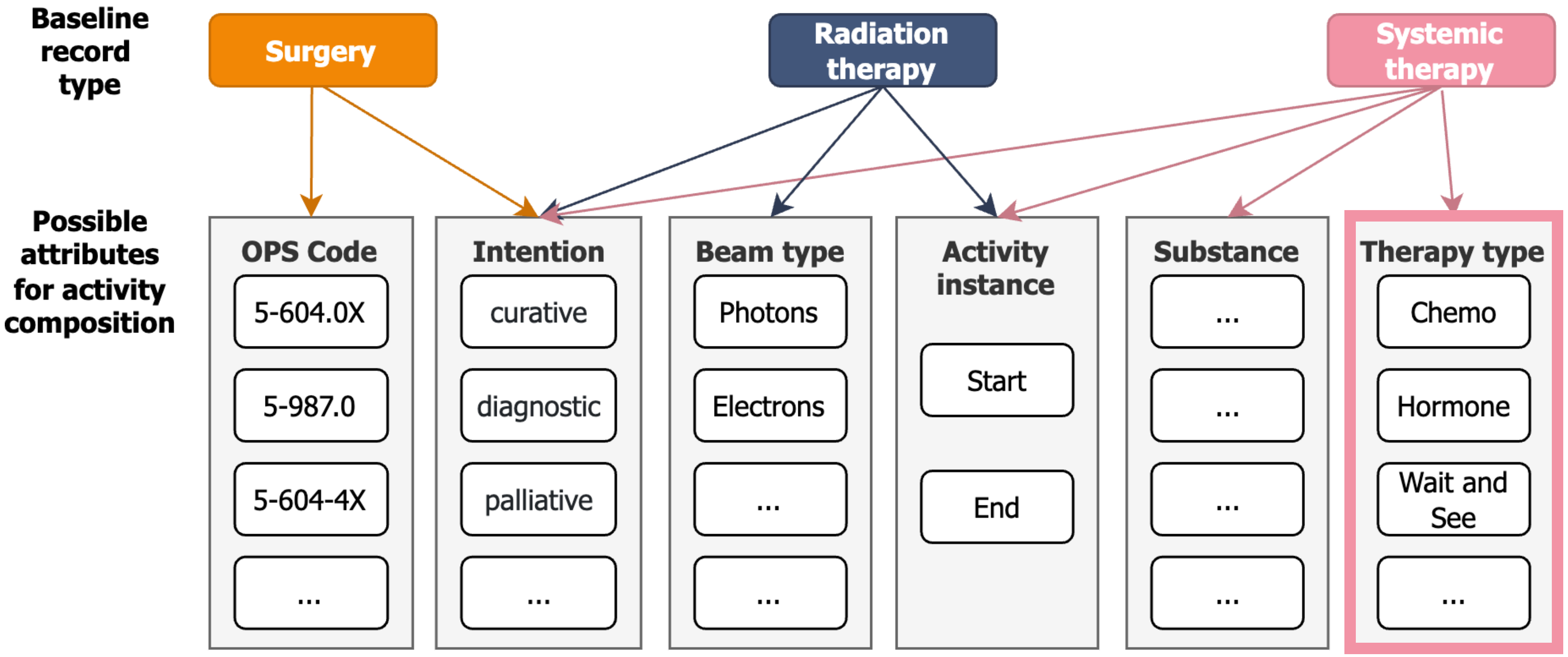
Grading
91,039

TMN
60,563

Record Types
- Master data
- Examinations
- Treatments
- Additional (attached) Records

d) Data Processing

- **Case ID:** Tumor ID
- **Timestamp:** Activity Date
- **Activity:** Record type (Diagnosis, Surgery...)
- Removal of implausible traces after quality check
- Final dataset: 13,069 complete cases between 2018 to 2023 (all events in the trace must be completely in the interval)

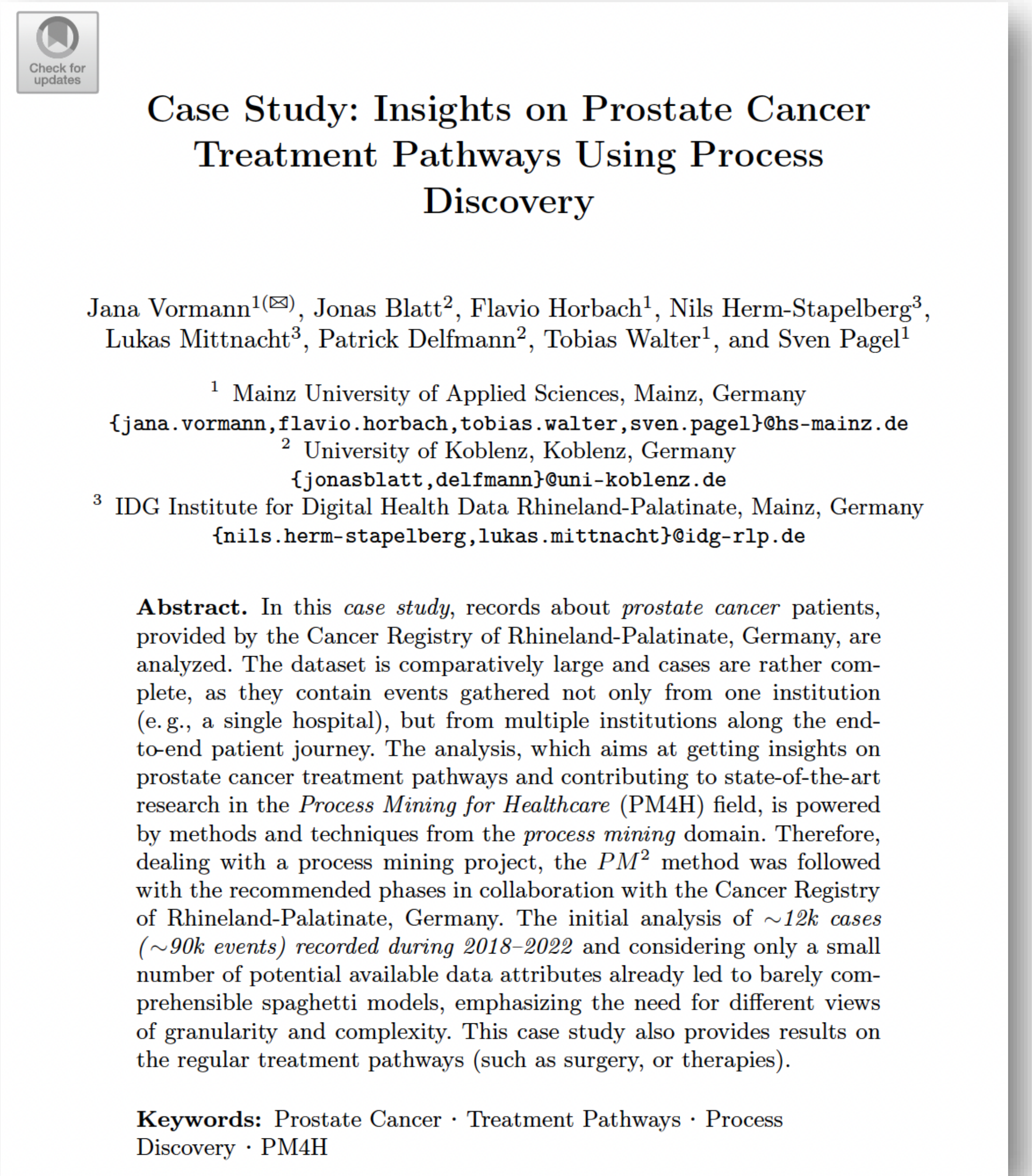


e) Process Discovery

For process discovery results see:

[Case Study: Insights on Prostate Cancer Treatment Pathways Using Process Discovery]

Jana Vormann, Jonas Blatt, Flavio Horbach, Nils Herm-Stapelberg, Lukas Mittnacht, Patrick Delfmann, Tobias Walter & Sven Pagel (2025)



Check for updates

Case Study: Insights on Prostate Cancer Treatment Pathways Using Process Discovery

Jana Vormann[1(✉)], Jonas Blatt[2], Flavio Horbach[1], Nils Herm-Stapelberg[3], Lukas Mittnacht[3], Patrick Delfmann[2], Tobias Walter[1], and Sven Pagel[1]

[1] Mainz University of Applied Sciences, Mainz, Germany
{jana.vormann,flavio.horbach,tobias.walter,sven.pagel}@hs-mainz.de
[2] University of Koblenz, Koblenz, Germany
{jonasblatt,delfmann}@uni-koblenz.de
[3] IDG Institute for Digital Health Data Rhineland-Palatinate, Mainz, Germany
{nils.herm-stapelberg,lukas.mittnacht}@idg-rlp.de

Abstract. In this *case study*, records about *prostate cancer* patients, provided by the Cancer Registry of Rhineland-Palatinate, Germany, are analyzed. The dataset is comparatively large and cases are rather complete, as they contain events gathered not only from one institution (e. g., a single hospital), but from multiple institutions along the end-to-end patient journey. The analysis, which aims at getting insights on prostate cancer treatment pathways and contributing to state-of-the-art research in the *Process Mining for Healthcare* (PM4H) field, is powered by methods and techniques from the *process mining* domain. Therefore, dealing with a process mining project, the PM^2 method was followed with the recommended phases in collaboration with the Cancer Registry of Rhineland-Palatinate, Germany. The initial analysis of *∼12k cases (∼90k events) recorded during 2018–2022* and considering only a small number of potential available data attributes already led to barely comprehensible spaghetti models, emphasizing the need for different views of granularity and complexity. This case study also provides results on the regular treatment pathways (such as surgery, or therapies).

Keywords: Prostate Cancer · Treatment Pathways · Process Discovery · PM4H

f) Process Improvement & Support

Improvement

- Organizational perspective to discover cross-organizational treatment patterns
- Conformance checking with clinical/treatment guidelines
- Performance analysis & advanced performance visualization

Future/Current research

- Appropriate pre-processing methods (e.g. abstraction, clustering)
- Treatment pathways in the course of time (changing treatment practices)
- Role of new technologies (e.g., surgery robots)
- Frequent treatment patterns
- Treatment predictions
- Treatment recommendations
- Consideration of other types of cancer

universität koblenz
Informatik

RheinlandPfalz
MINISTERIUM FÜR WISSENSCHAFT UND GESUNDHEIT

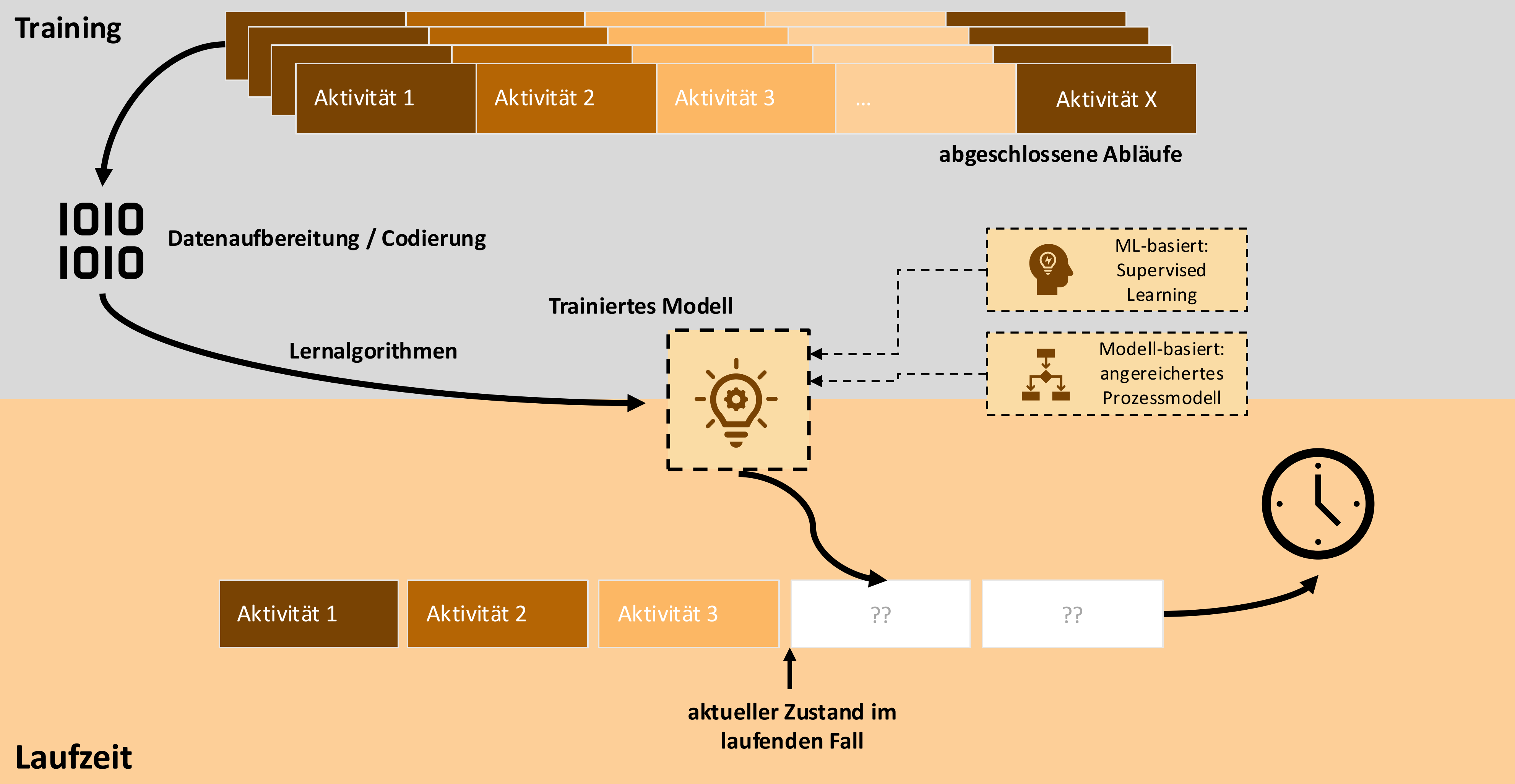
HOCHSCHULE MAINZ
UNIVERSITY OF APPLIED SCIENCES

4. Predictive Process Monitoring of Treatment Pathways

23/04/2025 13:05 – 13:35

a) Predictive Process Monitoring (Throwback)
b) Prediction Model Engineering
    1. Feature Engineering
    2. Prediction Problem
    3. Modell Architecture
c) Evaluation
d) Further Research

a) Predictive Process Monitoring (Throwback)



b1) Feature Engineering

Event Log

+

Categoric Features
ICD-Code
Residialstatus
OPS-Code
Intention
Tumorstatus
Stellung OP
Therapieart
Ende Grund
Applikationsart

+

Numeric Features
Duration
OPS-Version

+

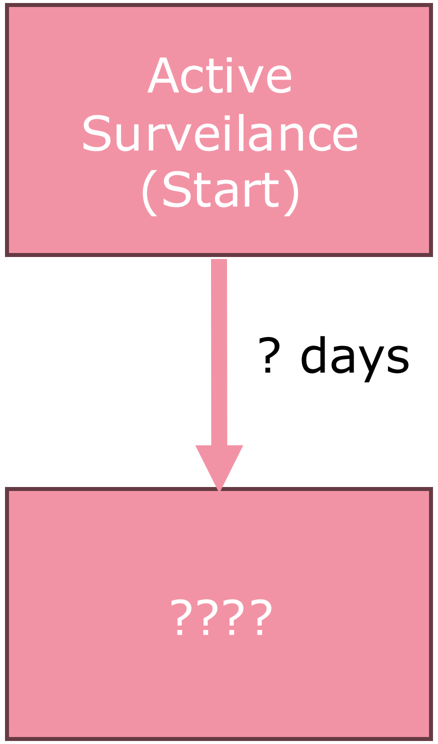
Trace Encoding

Diagnosis → Surgery → Radiation → Systemic Therapy

Prefix-Splits:
[Diagnosis] → pred *Surgery*
[Diagnosis, Surgery] → pred *Radiation*
[Diagnosis, Surgery, Radiation] → pred *Systemic*

b2) Prediction Problem

- ***Next Activity***
  - Predict the next activity based on previous events and additional features
  - ! Can be complicated because of lifelong cancer treatment / surveillance („cancer-free“ not „cured“)
  - ! First activity prediction can also be complicated due to no previous data



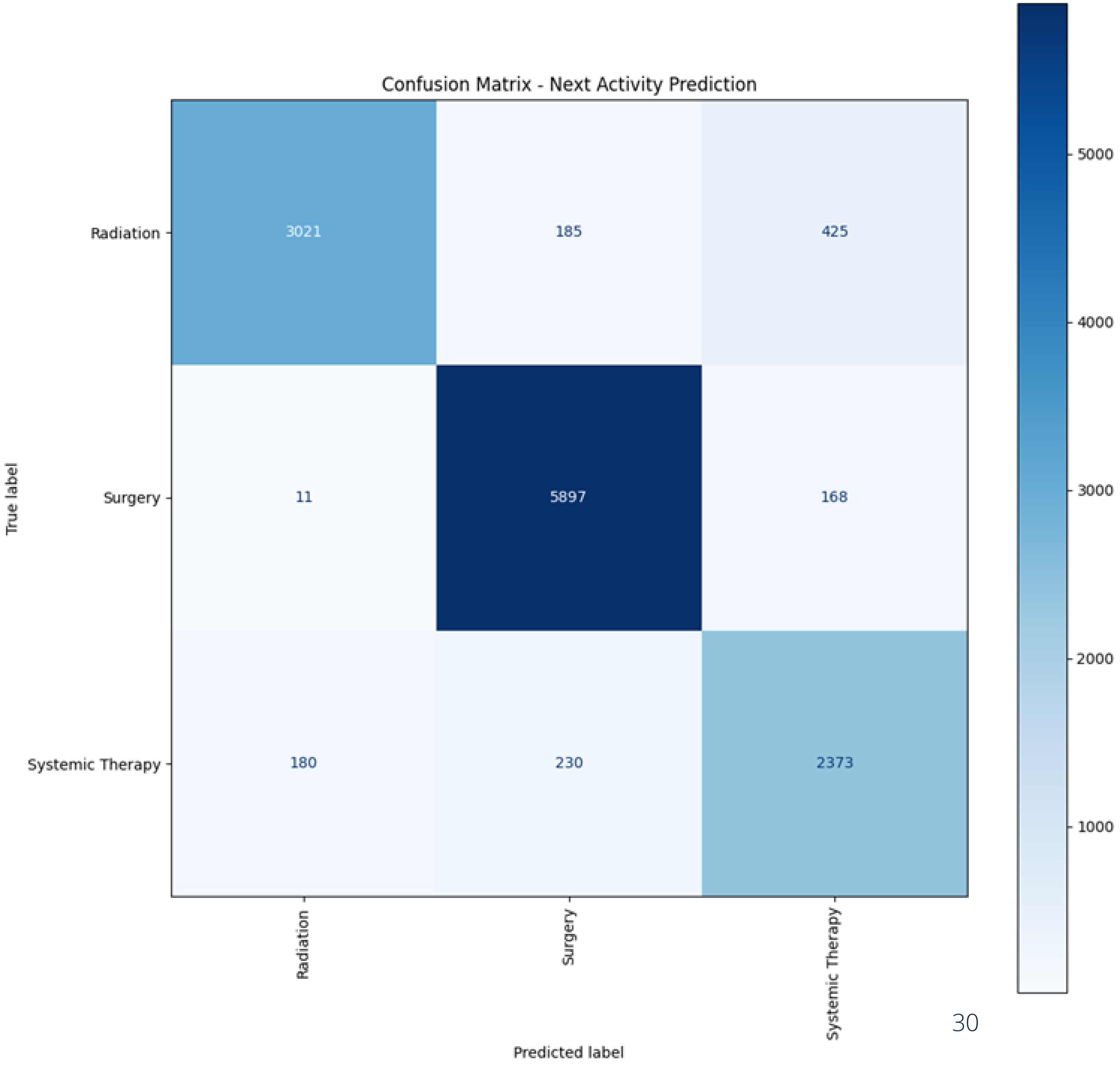
c) Evaluation: Synthetic prediction example

Index	Predicted	Actual
1 2 3 4 5	Surgery Surgery Surgery Radiation Radiation	Surgery Surgery Radiation Radiation Radiation
1 2 3	Systemic Therapy Systemic Therapy Systemic Therapy	Systemic Therapy Systemic Therapy Systemic Therapy
1 2 3 4 5	Systemic Therapy Tumor Conference Surgery Surgery Surgery	Systemic Therapy Diagnosis Surgery Surgery Surgery

0 is most of the times a Diagnosis

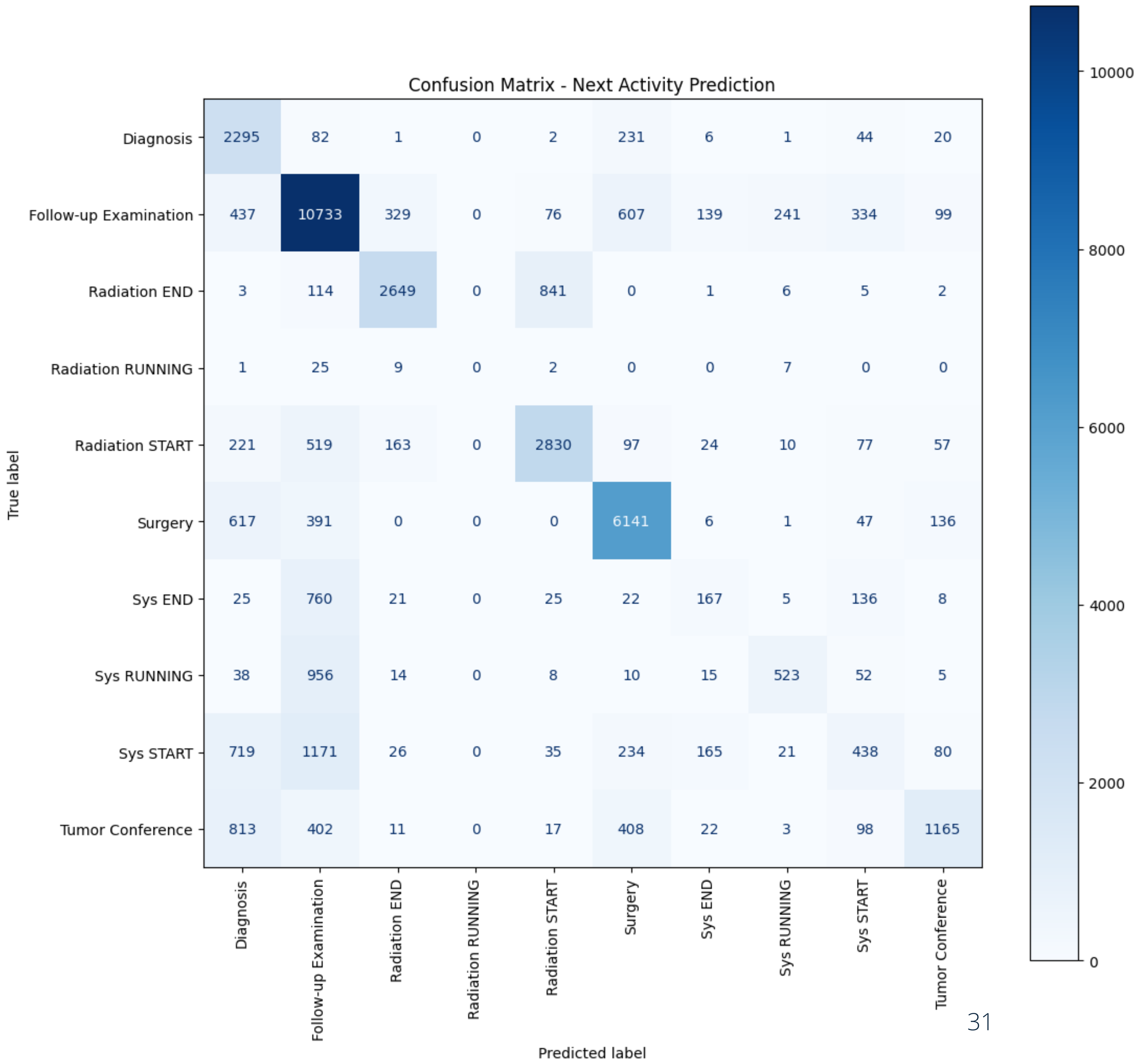
c) Evaluation: Treatment Model

- Accuracy: 0.905
- Only treatments
- No Start / End Split
- No split for Systemic therapy



c) Evaluation: Complete Start / End Model

- Accuracy: 0.61
- Complete activities
- Start / End / Still Running Split



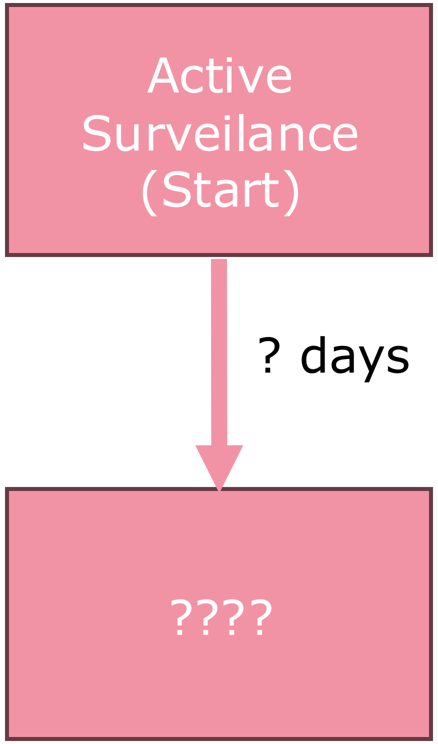
d) Further Research

Performance Aspects

- When will the patient receive the next treatment?
- Duration of the single treatment or entire path
- Probability of dropping out of therapy
- Probability of „success" of a path

Extension: Recommendations

- Recommend next treatment for „successful" therapy/path
- Recommend first treatment after Diagnosis Activity



Cancer Treatment Pathways: Caution with AI

At first glance "just" data, but behind all data lies the health of a person.

Always a human in the loop approach required.

Bildnachweis: DALL·E 3

5. Ausblick

23/04/2025 13:05 – 13:35

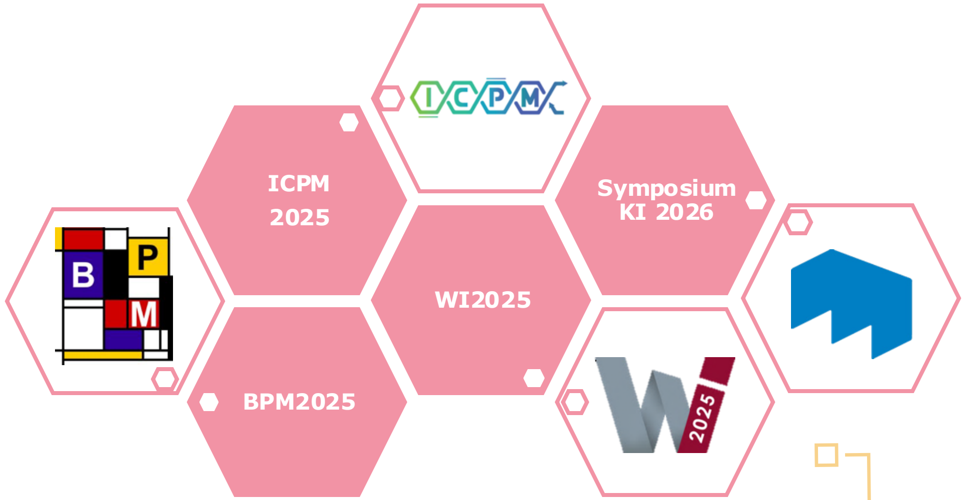
Stay tuned ….

Kooperationsmöglichkeiten:

- Welche Daten haben Sie?
- Eigenen sich auch Ihre Daten für (Predictive) Process Monitoring?
- Offen für neue Partner im Bereich Medizin (auch andere denkbar)

Verfolgen Sie unsere Projektfortschritte:

ai-dpa.hs-mainz.de



Zeit für Ihre Fragen

Kontakt

Hochschule Mainz

Prof. Dr. Sven Pagel
Mail: sven.pagel@hs-mainz.de

Prof. Dr. Tobias Walter
Mail: tobias.walter@hs-mainz.de

Jana Vormann M. Sc.
Mail: jana.vormann@hs-mainz.de

Flavio Horbach M. Sc.
Mail: flavio.horbach@hs-mainz.de

Literaturverzeichnis

- Appice, A., & Malerba, D. (2016). A Co-Training Strategy for Multiple View Clustering in Process Mining. IEEE Transactions on Services Computing, 9(6), 832–845. doi:10.1109/TSC.2015.2430327
- Chen, K., Abtahi, F., Carrero, J.J., Fernandez-Llatas, C., Seoane, F. (2023). Process mining and data mining applications in the domain of chronic diseases: A systematic review. Artificial Intelligence in Medicine 144, 102645. doi: 10.1016/j.artmed.2023.102645
- Dallagassa, M.R., Dos Santos Garcia, C., Scalabrin, E.E., Ioshii, S.O., Carvalho, D.R. (2022). Opportunities and challenges for applying process mining in healthcare: asystematic mapping study. Journal of Ambient Intelligence and Humanized Computing 13(1), 165–182. doi: 10.1007/s12652-021-02894-7
- Martin, N., Wittig, N., Munoz-Gama, J. (2022). Using Process Mining in Healthcare. In: Process Mining Handbook, LNBIP, vol. 448, pp. 416–444. Springer, Cham. doi: 10.1007/978-3-031-08848-3_14
- Munoz-Gama, J., Martin, N., Fernandez-Llatas, C., Johnson, O. A., Sepúlveda, M., Helm, E., Galvez-Yanjari, V., Rojas, E., Martinez-Millana, A., Aloini, D., Amantea, I. A., Andrews, R., Arias, M., Beerepoot, I., Benevento, E., Burattin, A., Capurro, D., Carmona, J., Comuzzi, M., … Zerbato, F. (2022). Process mining for healthcare: Characteristics and challenges. Journal of Biomedical Informatics, 127, 103994. doi: 10.1016/j.jbi.2022.103994
- Robert Koch-Institut, G.d.e.K.i.D. e.V. (2023): Krebs in Deutschland für 2019/2020. 14. Ausgabe. Robert Koch-Institut, Berlin.

Literaturverzeichnis

- Ronckers, C., Spix, C., Trübenbach, C., Katalinic, A., Christ, M., Cicero, A., Folk-erts, J., Hansmann, J., Kranzhöfer, K., Kunz, B., Manegold, K., Büschenfelde,U., Penzkofer, A., Vollmer, G., Weg-Remers, S., Barnes, B., Buttmann-Schweiger,N., Dahm, S., Franke, M., Schönfeld, I., Kraywinkel, K., Wienecke, A. (2023): Krebs in Deutschland für 2019/2020. Robert Koch-Institut and Gesellschaft der epidemi-ologischen Krebsregister in Deutschland e.V.. doi: 10.25646/11357
- Sánchez-Charles, D., Carmona, J., Muntés-Mulero, V., & Solé, M. (2018). Reducing Event Variability in Logs by Clustering of Word Embeddings. In E. Teniente & M. Weidlich (Hrsg.), Business Process Management Workshops BPM 2017 (Bd. 308, S. 191–203). Springer International Publishing. doi: 10.1007/978-3-319-74030-0_14
- Wang, P., Tan, W., Tang, A., & Hu, K. (2018). A Novel Trace Clustering Technique Based on Constrained Trace Alignment. In Q. Zu & B. Hu (Hrsg.), Human Centered Computing HCC 2017 (Bd. 10745, S. 53–63). Springer International Publishing. doi: 10.1007/978-3-319-74521-3_7
- Van Eck, M.L., Lu, X., Leemans, S.J.J., Van Der Aalst, W.M.P. (2015). PM2: A Process Mining Project Methodology. In: Advanced Information Systems Engineering. pp. 297–313. doi: 10.1007/978-3-319-19069-3_19
- Van der Aalst, W. (2016). Process Mining: Data Science in Action. Springer-Verlag, Berlin.
- Yin, R.K. (2014): Case Study Research: Design and Methods, 5th ed. SAGE Publications, Thousand Oaks, CA, USA.