Linking Data to Care: Mining the Pathways of Oncology

OP14

Alice Andalò, Lucia Bertoni, Filippo Merloni, Roberta Maltoni, Paolo De Angelis, Nicola Gentili
IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", Meldola, Forlì-Cesena, Italy

Objective

Process mining has emerged as a valuable approach for **analyzing real-world** healthcare **processes**, yet its application in oncology remains limited. We report the first **implementation** of process mining at the **IRCCS** Istituto Romagnolo per lo Studio dei Tumori (**IRST**) "Dino Amadori," an Italian cancer center, focusing on the care pathway of early-stage breast cancer patients.

Figure 1. Schematic representation of the process mining flow.

Methods

We analyzed the oncological pathways of **4.384** patients who underwent their first oncology consultation between **2015 and 2023** and were diagnosed with **breast cancer**. Administrative and clinical data were extracted and preprocessed to construct event logs, enabling process discovery with the **PM4Py** library. Stratification by cancer stage improved the clarity of the models; based on this, we selected **2.785** cases diagnosed at an early stage. Data-driven pathways were compared with the regional clinical guidelines to assess adherence and identify process deviations.
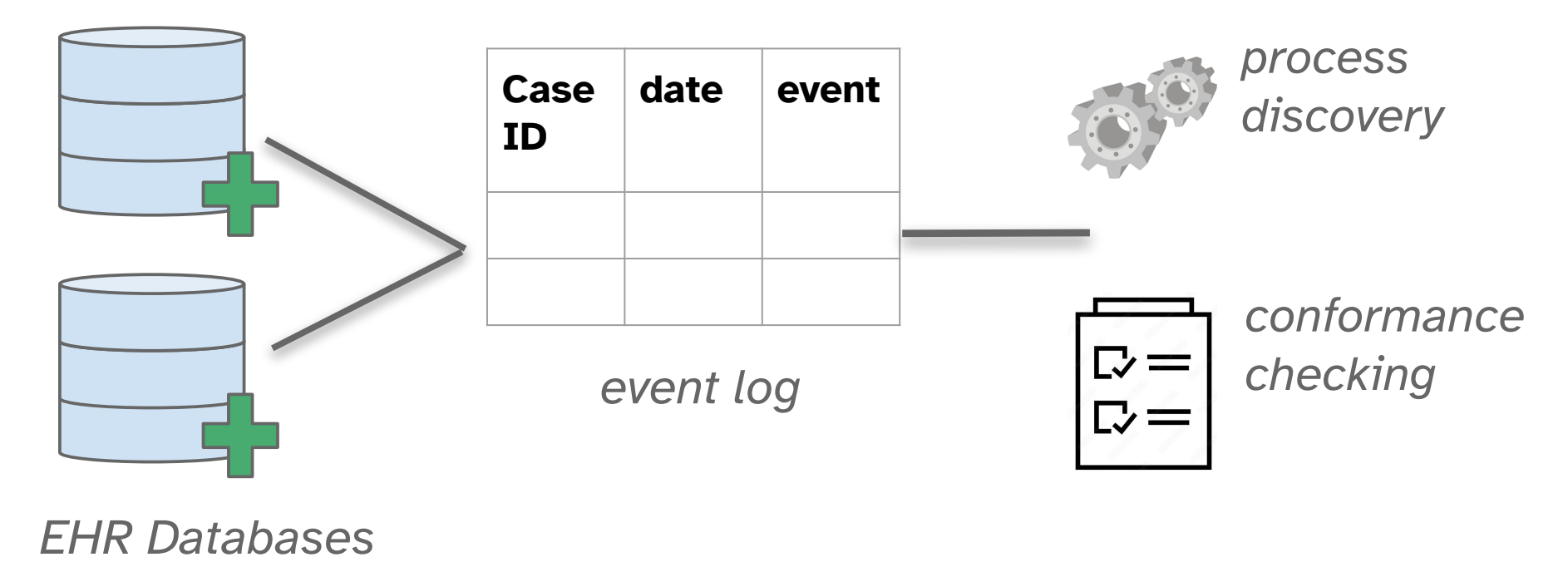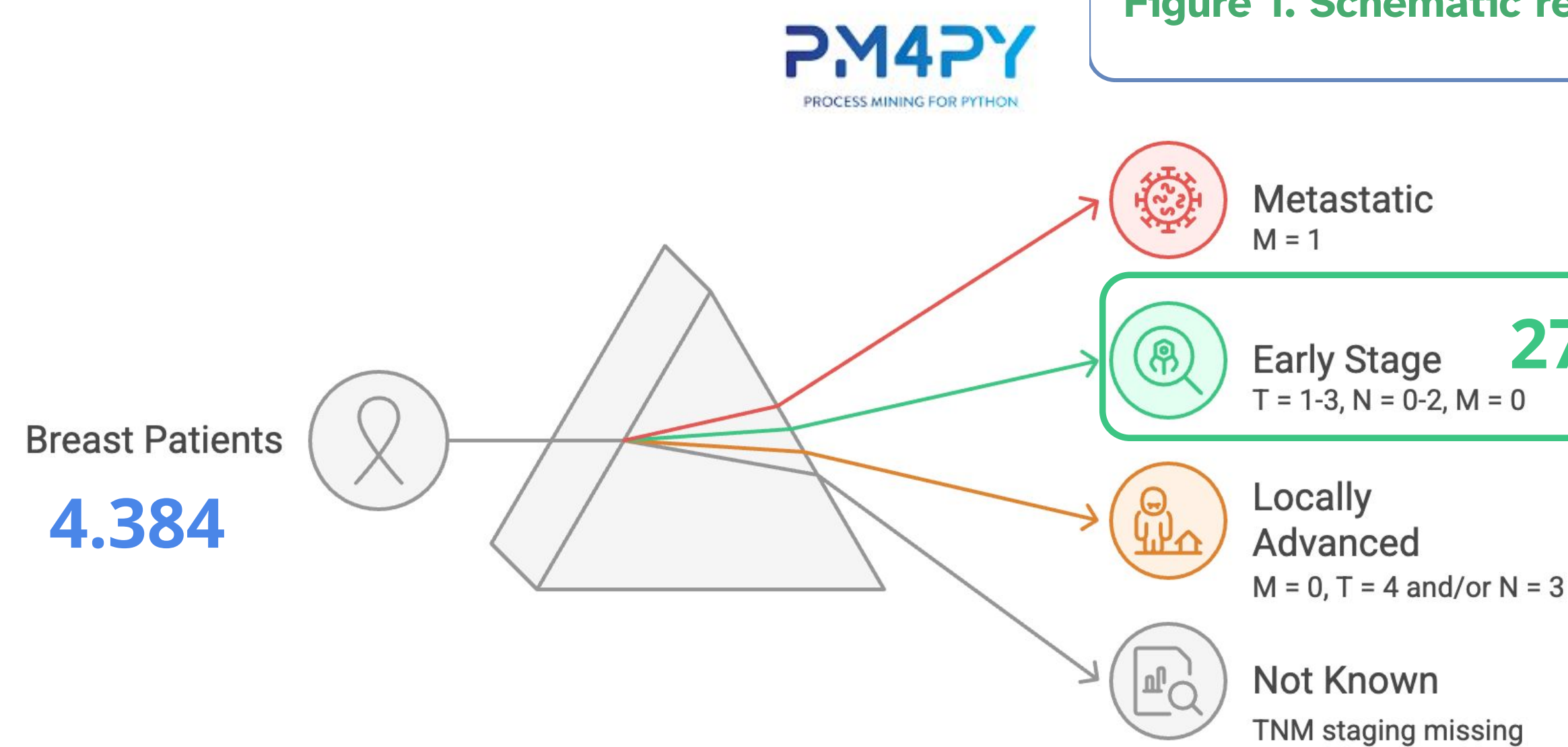

Figure 2. Distribution of breast cancer patients by disease stage. The cohort includes 4.384 breast cancer patients. Patients were classified according to TNM staging into four groups: metastatic (M = 1), early stage (T = 1–3, N = 0–2, M = 0), locally advanced (M = 0, T = 4 and/or N = 3), and not known (TNM staging missing). Among these, 2.785 patients were diagnosed at an early stage and selected for subsequent analyses.

Results & Discussion

concept:name	coverage_%
Diagnosis: Breast Cancer	100%
Initial visit-oncology	99.21%
Mastectomy or other surgical procedure	78.58%
Initial visit-radiotherapy	64.34%
Diagnostic without radiation	58.24%
Start RT	49.28%
End RT	49.18%
Laboratory	39.08%
Start Chemotherapy	35.44%
End Chemotherapy	35.27%
Diagnostic with radiation	31.16%
Biopsy	25.13%

Fig. 3. Table summarizing the clinical events and their percentage of coverage across the entire cohort of early-stage breast cancer patients. The coverage also reflects the overall quality of the data retrieved.

The resulting **process maps** provided a **clear visualization** of the main **care pathways**, particularly after refining the event logs through stratification and filtering techniques. The analysis confirmed alignment with core clinical milestones while also highlighting opportunities to optimize pathway adherence. A focused evaluation of time to chemotherapy (**TTC**) showed that approximately **84%** of patients initiated treatment within the clinically recommended **60-day window**.
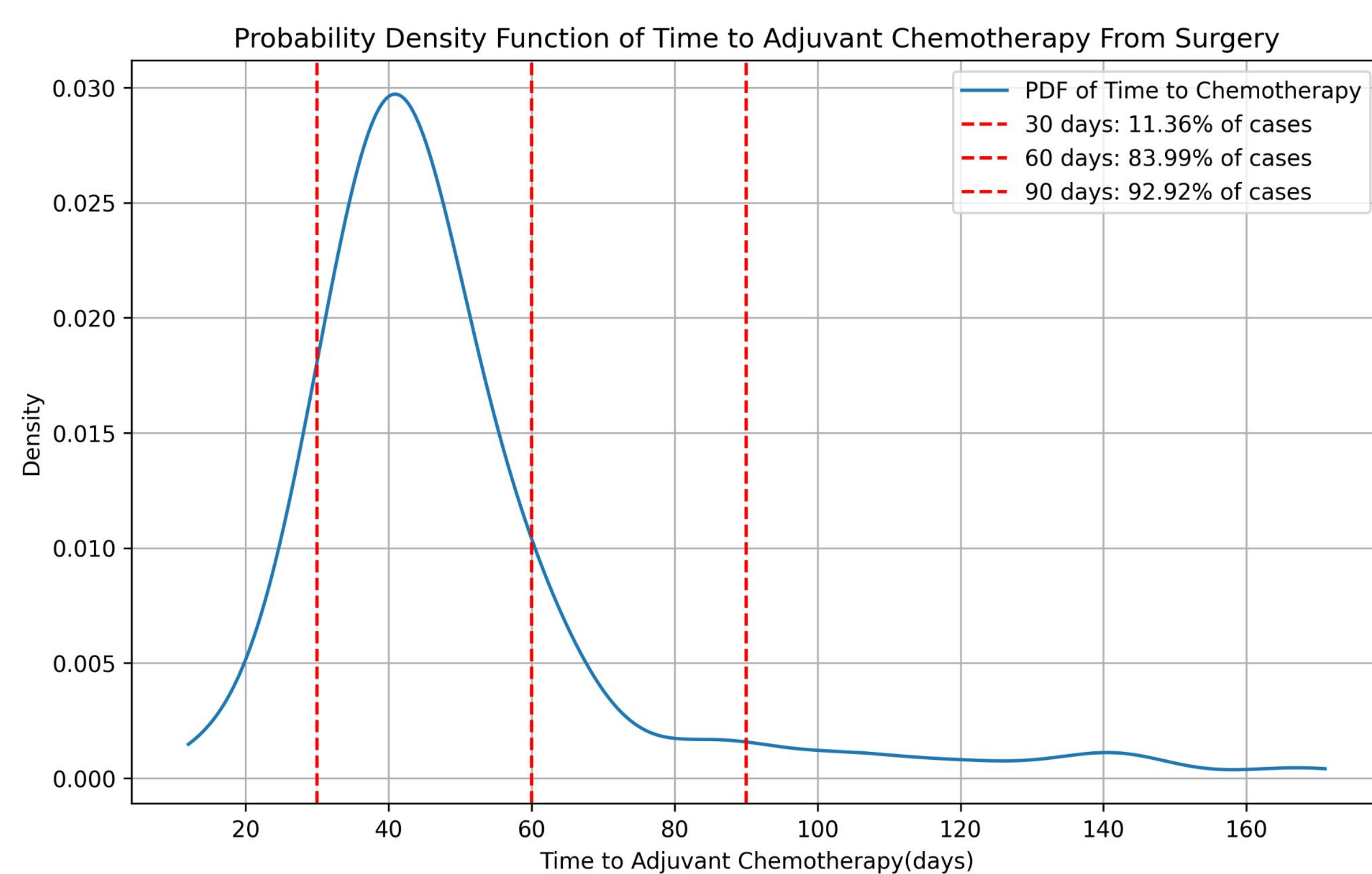

Fig. 4. Probability density function of the time distribution from surgery to the start of adjuvant chemotherapy, highlighting the critical thresholds (30, 60, and 90 days).
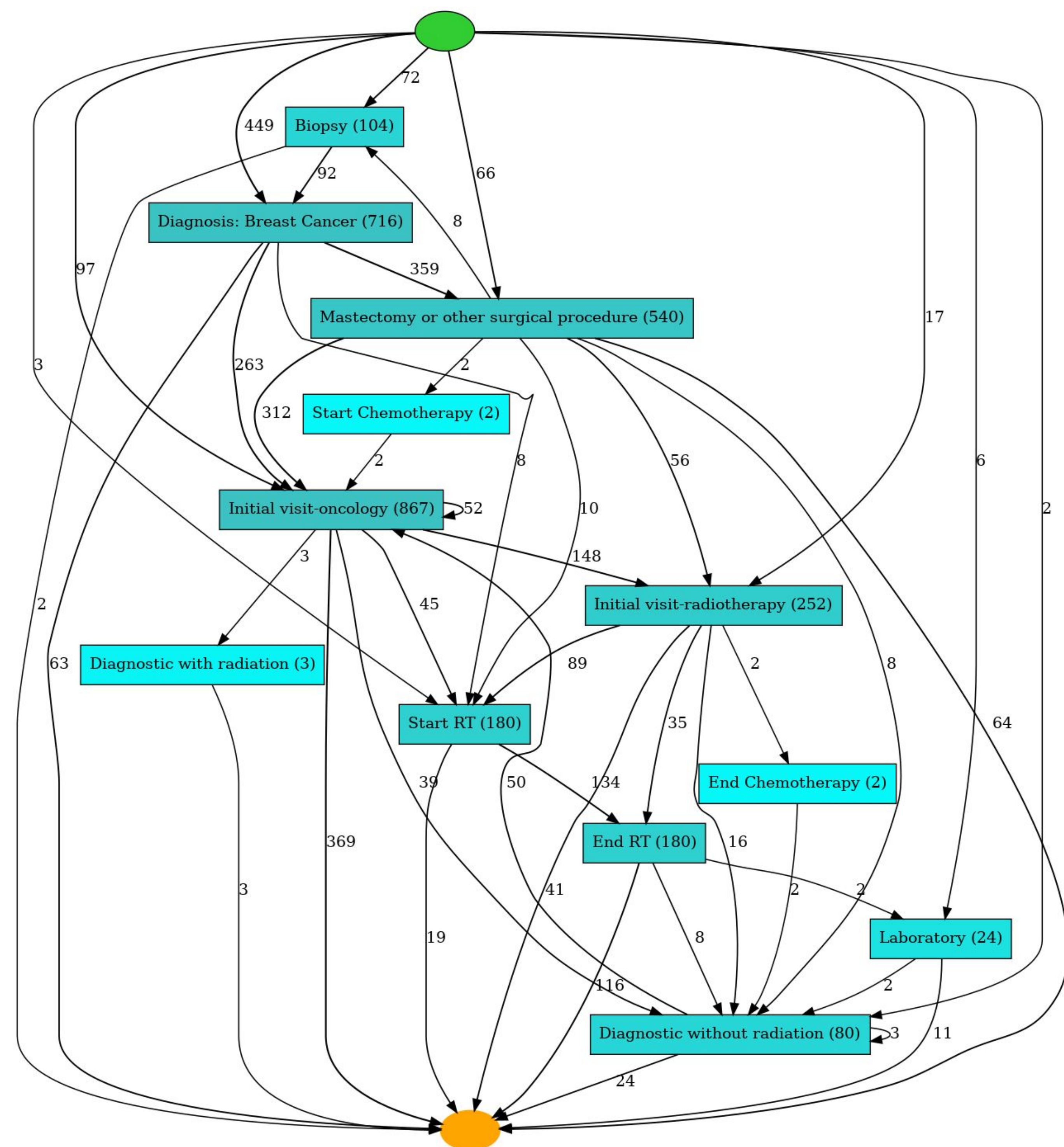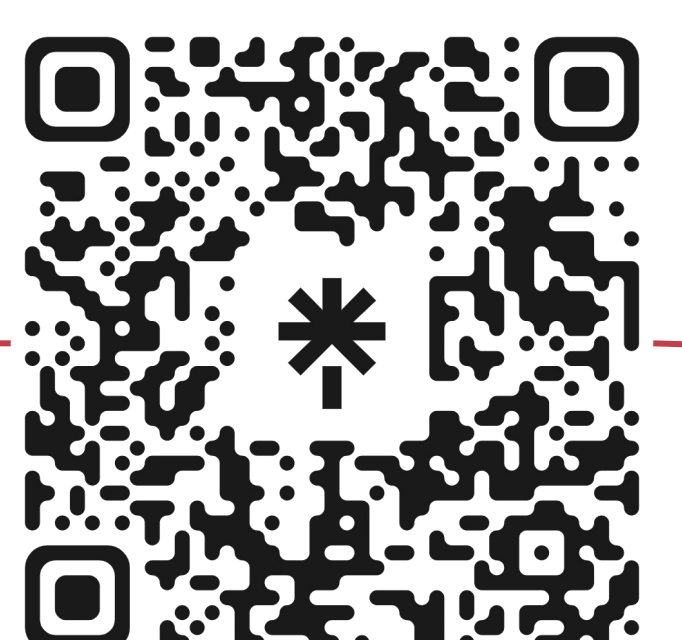

Fig. 5. Heuristic Net representing the pathway map for patients with early-stage breast cancer.

Throughout the project, **iterative feedback** from a **multidisciplinary** group of clinicians and hospital administrators played a pivotal role in interpreting results and validating model relevance. This collaboration enabled a more nuanced understanding of deviations from expected care pathways.

Conclusions

Our study demonstrates the **feasibility** of **process-oriented data science** in oncology. The approach offers a reproducible methodology for deriving actionable insights and supporting evidence-based improvements in complex healthcare settings.

Want to know more about us?

dataunit@irst.emr.it