Predictive Insights for Personalising Esophagogastric Cancer Treatment Process - A Case Study

Predictive Insights for Personalising Esophagogastric Cancer Treatment Process - A Case Study

Mozhgan Vazifehdoostirani[1], Andrei Buliga[2,3], Laura Genga[1], Rob Verhoeven[4,5,6], and Remco Dijkman[1]

[1] Eindhoven University of Technology, the Netherlands
[2] Free University of Bolzano, Italy
[3] Fondazione Bruno Kessler, Italy
[4] Netherlands Comprehensive Cancer Organisation (IKNL), the Netherlands
[5] Amsterdam UMC, location AMC, the Netherlands
[6] Cancer Center Amsterdam, Cancer Treatment and Quality of Life, the Netherlands

Abstract. For metastatic esophagogastric cancer (EGC), treatment aims to extend survival time. However, determining the best treatments for patients with EGC is challenging due to patients' variability. Personalised treatments supported by prediction models enable tailoring treatment process to individuals. Even so, traditional prediction models often neglect the interaction between treatments, limiting their utility in comprehensive planning. State-of-the-art Predictive Process Monitoring shows promising results in predicting the outcome of the treatment process but often lacks transparency. This paper investigates the potential of supporting healthcare experts in personalising the EGC treatment process, using eXplainable Predictive Process Monitoring methods. A real-world case study among 7,090 patients identifies expert needs for helpful explanations and discusses the capabilities and limitations of existing methods, suggesting future research directions. Our findings demonstrate high-quality explanations with strong *fidelity*, providing insights validated by expert knowledge. While the resulting explanations are not always actionable, experts acknowledged their value for exploratory analysis.

Keywords: Healthcare Processes · Explainable Predictive Process Monitoring · Process Pattern

1 Introduction

Stomach and esophageal cancer (combined as esophagogastric cancer, EGC) are in the top ten most common cancers worldwide [16]. The goal of palliative treatment procedures for many patients with advanced stages, including those with metastatic EGC, is to extend overall survival by managing symptoms and enhancing patients' quality of life [4]. However, there is no consensus on the best

treatment process. The high variability in disease aspects, patients' characteristics, desires, and dynamic conditions make it challenging for experts to determine the most effective treatments for an individual patient [15].

Prediction models can aid in personalising the treatment process and allow individualized decision-making [20]. By leveraging historical data, prediction models can identify patterns and correlations that may not be apparent to human practitioners [13]. However, traditional prediction methods often focus on the demographic and clinical characteristics of the patient and ignore the order in which the individual treatments are provided (which we refer to as the *treatment process*) [10, 15, 16]. As shown in [18], the impact of activities on the process outcome might vary depending on their position within the trace. Therefore, ignoring treatment orders limits the utility of previous prediction methods in comprehensive treatment planning. Additionally, considering the dynamic changes in patient conditions, practitioners require methods that enable them to monitor patient and disease progression and predict outcomes at different stages of the treatment process. Recent work in the field of Predictive Process Monitoring (PPM) provided ML-based prediction methods that consider the order in which individual treatments are provided for each patient [13]. These models have shown promise in effectively predicting the treatment process's outcome but often function as *black boxes*, offering little insight into how predictions are made or how to adjust treatment processes for a desired outcome.

This lack of transparency is a significant drawback when utilizing black box models in the healthcare domain [12]. Our healthcare partners emphasize that the prediction results must be interpretable for informed and data-driven decision-making. Furthermore, the explanations provided must be clinically sound and relevant. Recent approaches in eXplainable Predictive Process Monitoring (XPPM) aim to address these issues by promoting effective collaboration between human experts and prediction models [11].

In this paper, we aim to investigate the potential of providing practitioners with explanations of black-box prediction models to support personalising the EGC treatment process. To this end, we conducted a real-world case study to identify experts' needs for useful explanations and assess the capability of existing methods to meet these needs. Particularly, we explore the benefits and limitations of the PABLO (PAttern-based LOcal explanation) [3], a state-of-the-art XPPM technique, as the best-fitting method for this case study according to our literature review and interviews with healthcare experts.

The paper is structured as: Section 2 provides background, Section 3 outlines the methodology, Section 4 presents the results and discusses potentials and limitations, and Section 5 concludes the paper.

2 Background

PPM aims at predicting the unfolding of ongoing process executions leveraging machine learning or deep learning models trained on event logs storing past process executions. Within PPM, there are several prediction tasks, e.g., predicting

the time until the completion of an execution or the next activity to be executed. In this paper, we focus on outcome-based PPM, where we predict the outcome of ongoing trace executions. Recently, outcome-based PPM has seen growing use in healthcare, showing promising results according to various studies [1, 5, 13].

A pressing issue in PPM relates to the accuracy/explainability trade-off: while more complex models perform better in terms of accuracy, they often lack interpretability [11]. Therefore, prior PPM studies have utilized post-hoc eXplainable Artificial Intelligence techniques to elucidate these black-box models. Consequently, a new subfield known as eXplainable Predictive Process Monitoring (XPPM) has emerged [6]. XPPM techniques can be classified into two categories: local and global explanations. Local explainability emphasizes personalized interpretation, supporting experts in understanding *why a prediction was made for a specific trace (i.e., an "inquiry trace")*. Global explainability provides an overview of prediction models at a population level, supporting *focus on strategic decisions* [14, 18]. Since this paper focuses on personalising treatment processes for individual patients, we explore local explainability methods.

Previous research has shown that process analysts frequently seek actionable insights within specific cases to achieve desired outcomes, known as counterfactual explanations [7]. This task is challenging, as the counterfactuals must conform to process constraints and closely match the inquiry trace, enabling experts to achieve desired outcomes with minimal changes. Furthermore, previous research indicates that neglecting factual explanations and concentrating on counterfactuals hinders the discovery of truly predictive process patterns [3]. To the best of our knowledge, only two methods incorporate both factual and counterfactual process-aware local explanations: the Loreley [7] and PABLO frameworks [3]. However, the Loreley method focuses exclusively on static attributes of traces and lacks control-flow-aware explanations, which are essential for practitioners when patient characteristics cannot be modified. Conversely, PABLO allows adjustments in both control flow and static attributes, offering an explanation focusing on the adjustment of treatments. Thus, in this paper, we adopt an instantiation of the PABLO framework as a best-fitting method for the goal and characteristics of the presented case study.

3 Methodology

This case study has been conducted following the PM^2 framework [17], outlining six steps in a process mining project: *Planning*, *Extraction*, *Data Processing*, *Mining and Analysis*, *Evaluation* and *Process Improvement and Support*. Due to space constraints, we only discuss key activities performed at each stage for the operationalisation of the PM^2 framework in the following sections. Note that the final step is beyond the scope of this project.

3.1 Planning

The objective of the planning step is to set up the case study and to determine the research questions. The main goal of this case study is to investigate the

possibility of supporting healthcare professionals in personalising the treatment of EGC patients, leveraging prediction models and predictive insights.

Through interviews with healthcare practitioners, we identified two key limitations in deploying prediction models for treatment personalising. Since the best performing models are often black boxes, they find it essential to 1) understand the driving factors behind predictions (aka. factual explanations) and 2) learn possible factors to refine treatment plans when faced with undesirable outcomes (aka. counterfactual explanations).

Thus, we define our primary research question in this case study as: *How can we support healthcare experts with explanations of prediction models that guide them on which factors to change or retain for improved treatment outcomes?*

3.2 Data extraction and preprocessing

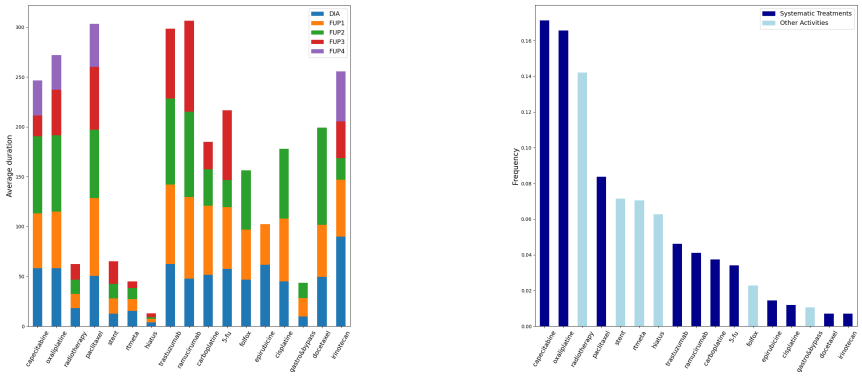
The dataset is provided by the Netherlands Cancer Registry, maintained by the Netherlands Comprehensive Cancer Organisation[1]. The dataset contains 7,090 metastatic patients between 2015-2021. We extracted data from all treatments performed on each patient after being diagnosed with EGC, including the timestamps, indicating their orders and duration. Additionally, we extracted 15 relevant patient attributes in outcome prediction reported in previous studies [16].

Data preprocessing was conducted with the guidance of healthcare professionals to ensure clinical relevance and accuracy. As part of this process, we removed 1,419 patients who had received only one treatment, focusing the analysis on those with more complex treatment histories. Additionally, we removed 553 cases where there were logging errors (e.g., patients for which the survival time was unknown) and exceptional cases, like patients who received one or multiple treatment(s) abroad. Similarly, 189 patients with too deteriorated health are removed from the dataset, as they are not fit enough to receive any treatment or to live more than 30 days. In the end, we concentrated solely on 3,217 deceased patients with observed complete survival times.

The initial log contained 161 distinct treatment codes. However, based on expert knowledge, multiple treatment codes could be grouped under common categories, reducing the complexity of the analysis. For example, 8 different types of Radiotherapy codes are relabelled to Radiotherapy. After adjusting the granularity level of the dataset, we also removed 170 patients who received a very rare treatment with lower than 0.5% frequency. The final dataset includes 3,047 patients exhibiting 807 process variants with 17 unique treatment codes.

The treatment process is divided into five episodes by experts, each ending with disease progression. We imputed the missed duration of treatments for 9% of events based on their average duration in the corresponding episode. As depicted in Figure 1a, some treatment durations vary across different episodes. We used treatment duration as a dynamic attribute for prediction models.

[1] The Dutch Central Committee on Research Involving Human Subjects deemed ethics approval unnecessary. The Dutch Upper-GI Cancer Group and the Netherlands Cancer Registry's Privacy Review Board approved the use of anonymous data.



(a) The average duration of each treatment in different episodes

(b) The distribution of systematic and non-systematic treatments

Fig. 1: Frequency and duration of each treatment after preprocessing

Eventually, we devised two scenarios based on experts' knowledge to generate life expectancy binary labels for each patient. This step is done to make the dataset compatible with existing methods, which are mainly designed for binary outcome prediction [9]. In the first scenario (L1), recorded survival times (days survived after completing the last treatment) were categorized as "Low" (162 days or fewer) or "High" (more than 162 days) based on expert knowledge, with 24% of cases labelled as "Low". In the second scenario (L2), the threshold for "Low" or "High" life expectancy varied by treatment type. Patients receiving any *systematic treatments* were classified as "Low" if their survival time was under 120 days, while those not receiving systematic treatment were classified as "Low" if their survival time was under 60 days. In this scenario, only 16% of cases were labelled as "Low". The frequency of each treatment is depicted in Figure 1b, with systematic treatments highlighted in dark blue and other treatments shown in light blue. We will report and analyze the results of two labelling strategies in the following sections.

3.3 Mining and analysis

Since the main research question requires a method to generate both factual and counterfactual explanations, we employ the PABLO framework (see Figure 2). In the following, we discuss the adoption of the PABLO framework for the EGC dataset according to our healthcare experts' desires and needs:

Predictive model training. We implemented the XGboost model as the black box model since it shows promising performance in previous studies [13].

To build the XGboost model, we need to encode each prefix trace in a format compatible with the model. For example, each prefix trace of length m like $\sigma =< e_1, \ldots, e_m >$ has to be represented through a feature vector F. In this case study, we used three encoding methods from PPM literature:

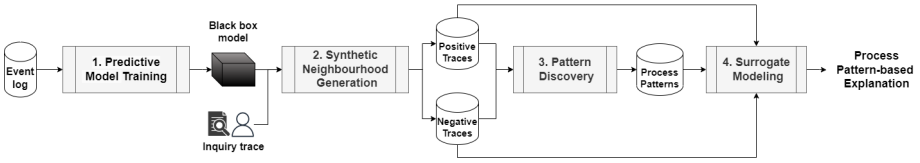


Fig. 2: The pattern-based local explanation (PABLO) framework

- In Simple-Index encoding (SI) [8], each feature corresponds to a position in the trace, with possible values being event classes, resulting in $F =< a_1, \ldots, a_m >$, where a_j is the event class a at position j.
- Simple-Trace-Index encoding (STI) [2] includes both dynamic sequence information and static trace attributes defined as $F =< s_1, \ldots, s_u, a_1, \ldots, a_m >$, where each s_u is u^{th} static feature.
- In Complex-Index encoding (CI) [8], the event classes and both static and dynamic data attributes of a trace are encoded in the vector. The resulting feature vector is $F =< s_1, \ldots, s_u, a_1, \ldots, a_m, h_1^1, \ldots, h_m^1, h_1^r, \ldots, h_m^r >$, where each h_m^r is a dynamic feature, corresponding to an event attribute.

We found the optimal XGBoost model setup for each encoding method on different prefix lengths via hyperparameter optimization using the Tree Parzen Estimator (TPE), maximizing the Area Under the Curve (AUC).

Synthetic neighbourhood generation. The inputs of the second component of the PABLO framework are a trained black-box predictive model and the inquiry trace as an encoded vector. The goal of this step is to generate a neighbourhood of similar cases with the same (aka, negative) and opposite (aka, positive) predicted outcomes for inquiry trace.

We used the Genetic Algorithm (GA) from [3] to generate the synthetic neighbourhood. Upon initial experiments, we found that the GA required adjustment due to the EGC dataset's unique characteristics, where longer traces often correspond to longer lifetimes (desired outcomes). As a result, the GA frequently produced impractically long counterfactual traces (in about 20% of synthetic neighbourhoods) rather than reordering treatments within shorter traces. To address this, we adjusted the GA to prioritize assigning activities to lower positions in subsequent iterations. This adjustment helps prevent the creation of lengthy traces when the inquiry trace is short. We employed two strategies to generate the synthetic neighbourhood, for different analysis purposes:

S1: In the first scenario, we synthetically generate traces with the same attributes as the inquiry case but alter the sequence of treatments to provide an actionable explanation where patient attribute changes are clinically unfeasible.

S2: In the second scenario, both treatment sequences and attributes are modified to generate the synthetic neighbourhood. Though patient attribute changes may not be clinically actionable, experts value this for exploratory analysis.

Process pattern discovery. We employed IMPresseD [19] as recommended in [3] to discover outcome-oriented control-flow patterns. IMPresseD adopts a

multi-objective optimization approach to discover patterns balancing between multiple interest functions [19]. We used *frequency*, *outcome*, and *likelihood* interest functions, introduced in [3]. We aim to discover patterns with high frequency and correlation with the outcome, which are obtained from traces with a high likelihood of belonging to the predicted outcome. Additionally, as experts are concerned about confounding variables, we used *case distance* interest function introduced in [19] as the fourth interest function. This metric ensures that a process pattern is genuinely discriminative by requiring a minimal distance between the initial status (attributes) of patients who received the treatment pattern and those who did not. If this distance is too large, the observed outcomes may be influenced by differences in patient characteristics rather than the pattern itself.

Surrogate modeling. A Decision Tree (DT) is used as a surrogate model to allow us to discover a rule-based explanation using the synthetically generated neighbourhood and discovered patterns from the previous step [3]. We have also found the optimal parameter for the DT using TPE, maximizing AUC.

3.4 Evaluation

The purpose of the evaluation stage is to use the analysis results to come up with ideas that help the project's goal. To this end, we first compare the performance of the prediction models using different encoding methods w.r.t AUC. Then, to measure the quality of the generated neighbourhood, we use the following measures introduced in [2]. The distance metric (DIST) calculates the average distance between an inquiry trace and generated traces. The diversity (DIV) measures the average pairwise distance among traces in the generated neighbourhood. The implausibility metric (IMPLAUS) evaluates the distance of synthetic traces from the reference population. The sparsity (SPARS) quantifies the average number of feature changes in each synthetic trace relative to the inquiry trace. Lastly, the conformance score (CONF) indicates the ratio of synthetic traces satisfying all process DECLARE constraints, which were discovered from the traces in the training set, with a support of 50%. Due to high variability in the log, we could not discover DECLARE constraints with higher supports.

To assess the quality of the final explanation, we measure the faithfulness of the provided explanation to the original black box model using the Local Fidelity (LF) measure [3]. The higher LF value indicates the higher capability of the surrogate model to mimic the behaviour of the black box model.

Additionally, we discuss the provided explanations with our healthcare partners to assess their validity and usefulness. The goal of this analysis is to determine whether the provided explanation aligns with expert knowledge. Moreover, we are interested in identifying the challenges and limitations of using the provided explanation in practice by communicating the final results to experts.

4 Evaluation results

Table 1 shows the average values for each measurement from Section 3.4 across different prefix lengths, while Figure 3 illustrates LF values by prefix length.

Table 1: Evaluation metrics for different labelling and encoding scenarios

labelling	Encoding	AUC	NG	LF	DIS	DIV	CONF	IMP	SPAR
L1	SI	74.12%	S1	78.12%	0.868	0.873	0.682	0.853	2.011
	STI	76.30%	S1	77.42%	0.531	0.262	0.913	1.779	2.321
			S2	87.07%	0.826	0.932	0.956	3.245	2.965
	CI	87.05%	S1	91.01%	0.911	0.104	0.875	14.04	5.584
			S2	78.40%	1.272	1.302	0.944	12.56	6.536
L2	SI	71.29%	S1	74.84%	0.852	0.858	0.844	0.600	1.714
	STI	70.43%	S1	78.08%	0.511	0.295	0.842	1.860	2.252
			S2	92.21%	0.862	0.926	0.806	1.724	3.977
	CI	85.48%	S1	82.54%	0.767	0.096	0.843	13.55	5.347
			S2	83.20%	0.917	1.021	0.919	4.795	3.690

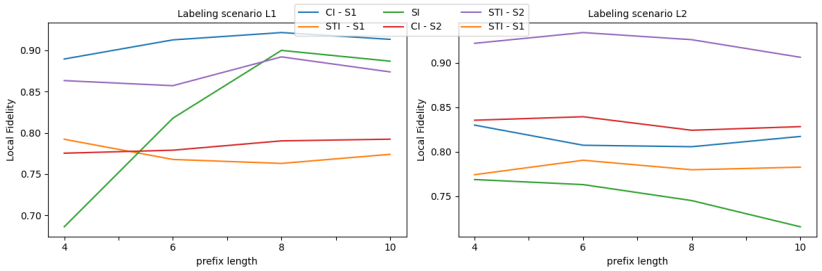


Fig. 3: Local fidelity over different prefix lengths

An initial observation from Table 1 is that CI encoding outperforms other methods in AUC due to its comprehensive attribute inclusion. The higher AUC suggests that the model is more accurate in distinguishing between patients who will benefit from certain treatments and those who will not, which is crucial for effective clinical decision-making. STI encoding's performance is close to SI encoding because STI only adds the initial state of patient attributes as static features. In contrast, CI encoding incorporates patient attributes as dynamic features and encodes all changes that occurred in the attributes through the process, leading to more precise predictions. Since the second labelling scenario (L2) depends on one group of treatments (systematic treatments) besides the remaining lifetime, the SI encoding is even slightly surpassing STI.

The first notable insight from analyzing the quality of the generated neighbourhood is a relatively high CONF measure in all scenarios (except for SI encoding in L1 labelling). Despite the 50% support for DECLARE constraints, which led to the discovery of constraints not present in all cases, GA was still able to generate highly conforming traces.

Inspecting other quality metrics for the synthetic neighbourhood, it appears that the GA method struggled to produce plausible traces using CI encoding. Despite relatively high LF values with CI encoding, the generated neighbourhood exhibited the highest IMP values, indicating a significant distance between the generated and original traces. The IMP is particularly higher for S1, where

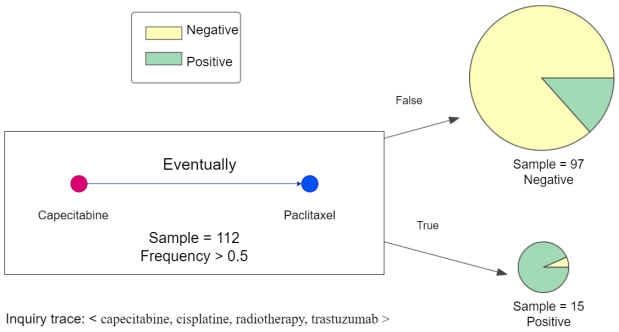


Fig. 4: Explanation example obtained from SI encoding for labelling scenario L1

modifications are restricted to the control flow. Given that the average trace lengths are generally quite short (4 treatments), there are limited possibilities for altering the control flow, resulting in the highest IMP and lowest DIV compared to other scenarios. In contrast, in S2, where all attributes can be altered, DIV increases, but the IMP and SPAR values remain high. This indicates that the distance to the original traces is still significant (high IMP), with many changes applied to the inquiry trace to generate the neighbourhood (high SPAR).

Overall, CI encoding provides more precise predictions and high-fidelity explanations. In contrast, STI and SI encoding methods deliver more valid and less intrusive explanations, thanks to the high-quality generated neighbourhood. These findings highlight the need for further research into enhancing neighbourhood generation using dynamic attributes.

To further explore the local fidelity results, we direct your attention to Figure 3. In L1 labelling, the combination of *CI - S1* consistently shows high fidelity, while *STI - S1* exhibits the lowest fidelity across all prefix lengths. Conversely, in L2 labelling, *STI - S2* outperforms other methods. In the L2, the rankings of the methods stay consistent across all prefix lengths. However, for L1, *SI* encoding improves significantly, going from the lowest LF at a prefix length of 4 to the second highest LF at a prefix of 10.

These contradictory behaviours between two labelling scenarios highlight the importance of defining the right outcome for a practical problem. It's worth mentioning that defining binary outcomes was challenging for experts because the final goal of the treatment process for EGC patients is to extend patients' remaining lifetime, which is, in essence, a regression problem. However, there is a lack of explanation methods designed specifically for regression problems [9], particularly when it comes to combined factual and counterfactual local explanations as needed for treatment process personalising.

4.1 Discussion

To discuss the potential and meaningfulness of the obtained explanations with healthcare practitioners, we presented explanations resulting from SI and STI

encoding, which offer higher neighbourhood quality, through a workshop. In the following, we highlight two interesting examples identified in the expert review.

Figure 4 presents the explanation obtained from a patient (inquiry trace) who received four treatments shown in the figure. DT nodes represent discovered process patterns and their frequencies, edges indicate whether frequency conditions are met, and the DT leaves show the predicted outcomes for those paths. The explanation suggests that administering *paclitaxel* eventually after *capecitabine* could shift the outcome from low to high remaining life class. The expert confirmed this explanation as a clinically sound counterfactual, having observed longer survival in patients receiving these two systematic treatments. However, if the patient is not fit enough for the second treatment, this counterfactual is not actionable. Due to SI encoding limitation, which ignores case attributes, the case condition is not presented in the final explanation.

Additionally, it would be valuable to know more details about the counterfactual treatment pattern, such as the time interval between these two treatments in the process pattern, especially when we have an eventual relation. A promising direction for future work is incorporating event data, such as timestamps, to enhance the discovery of predictive process patterns.

Another explanation example is depicted in Figure 5. With STI encoding, DT nodes may include patient attributes and their conditions or discovered patterns, similar to the previous example. The inquiry patient, in this case, received only two treatments (*hiatus* and *stent*), likely due to a bad initial health condition, as indicated by a health performance score of 3 (very poor condition). The short inquiry trace resulted in similarly short neighbourhood traces, limiting the discovery of larger process patterns. Consequently, the explanation often relies on a single treatment rather than a combination. The healthcare expert validated the relevance of the treatments provided as counterfactual for this case.

The counterfactual suggests that *Trastuzumab* could extend the patient's lifetime. If *Trastuzumab* is not an option, *Paclitaxel* with ct=1 may be beneficial. If neither treatment is feasible, setting perf_stat (health performance score) to 0 (very healthy) may shift the outcome favourably. Improving health performance makes sense as the last option, however this counterfactual is not actionable for metastatic patients. This explanation is interesting and could be relevant for others where improvements in health performance through exercise and nutrition are possible. Nevertheless, such explanations can validate whether the black box model has learned from the right and meaningful features and patterns and can assist physicians in trusting the prediction made by the black box model.

While our examination of explanations with healthcare experts reveals promising insights confirmed by expert knowledge, these explanations are not always actionable in practice. For example, the discussed counterfactual scenarios highlight important considerations but also underscore the need for additional information to fully assess their applicability. Specific details, such as treatment intervals, are crucial for making these explanations more actionable.

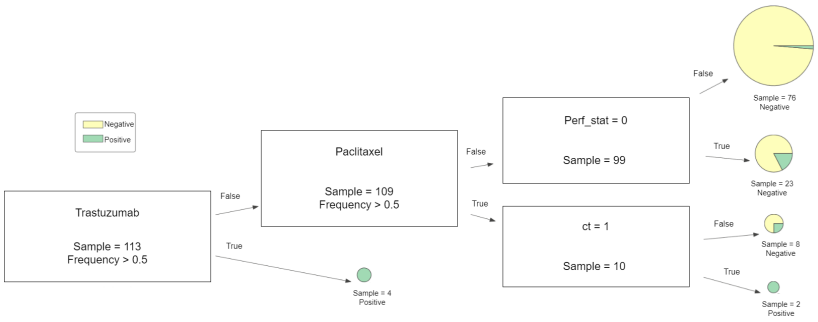


Fig. 5: Explanation example obtained from STI encoding for L2 and S2 scenarios

5 Conclusion

In conclusion, our investigation into the potential of XPPM methods for supporting EGC treatment personalisation highlighted the significant promise of these methods for clinical decision-making. Our real-world case study, focusing on PABLO framework, revealed high fidelity of these explanations (nearly 90%), conforming with expert knowledge. However, further analysis showed that the provided explanations are not always actionable in real-world scenarios. This limitation underscores the need for further refinement to ensure that the explanation can be effectively translated into clinical practice. Nevertheless, our healthcare partners expressed a strong interest in exploratory explanations, which can help to discover existing relationships and support a deeper understanding of black-box models. Also, given the evolving nature of dynamic attributes, making connections between dynamic attributes and process patterns could offer richer explanations and is a promising area for future work.

References

1. Jorg WR Bekelaar, Jolanda J Luime, and Renata M de Carvalho. Predicting patient care acuity: An lstm approach for days-to-day prediction. In *International Conference on Process Mining*, pages 378–390. Springer, 2022.
2. Andrei Buliga, Chiara Di Francescomarino, Chiara Ghidini, and Fabrizio Maria Maggi. Counterfactuals and ways to build them: Evaluating approaches in predictive process monitoring. In *International Conference on Advanced Information Systems Engineering*, pages 558–574. Springer, 2023.
3. Andrei Buliga et al. Uncovering patterns for local explanations in outcome-based predictive process monitoring. In *International Conference on Business Process Management*, pages 363–380. Springer, 2024.
4. Willemieke PM Dijksterhuis et al. Hospital volume and beyond first-line palliative systemic treatment in metastatic oesophagogastric adenocarcinoma: a population-based study. *European Journal of Cancer*, 139:107–118, 2020.
5. Annika L Dubbeldam, István Ketykó, Renata M de Carvalho, and Felix Mannhardt. Early predicting the need for aftercare based on patients events from

the first hours of stay–a case study. In *International Conference on Process Mining*, pages 366–377. Springer, 2022.
6. Ghada Elkhawaga, Mervat Abu-Elkheir, and Manfred Reichert. Explainability of predictive process monitoring results: Can you see my data issues? *Applied Sciences*, 12(16), 2022.
7. Tsung-Hao Huang, Andreas Metzger, and Klaus Pohl. Counterfactual explanations for predictive business process monitoring. In *EMCIS 2021, Proc.*, volume 437, pages 399–413, 2021.
8. Anna Leontjeva, Raffaele Conforti, Chiara Di Francescomarino, Marlon Dumas, and Fabrizio Maria Maggi. Complex symbolic sequence encodings for predictive monitoring of business processes. In *BPM 2015*, volume 9253, pages 297–313, 2015.
9. Simon Letzgus et al. Toward explainable artificial intelligence for regression models: A methodological perspective. *IEEE Signal Processing Magazine*, 39:40–58, 2022.
10. Xiaoyuan Ma, Eric Pierce, Harsh Anand, Natalie Aviles, Paul Kunk, and Negin Alemazkoor. Early prediction of response to palliative chemotherapy in patients with stage-iv gastric and esophageal cancer. *Bmc Cancer*, 23(1):910, 2023.
11. Nijat Mehdiyev and Peter Fettke. Explainable artificial intelligence for process mining: A general overview and application of a novel local explanation approach for predictive process monitoring. *Interpretable artificial intelligence: A perspective of granular computing*, pages 1–28, 2021.
12. Jorge Munoz-Gama et al. Process mining for healthcare: Characteristics and challenges. *Journal of Biomedical Informatics*, 127:103994, 2022.
13. Pam Pijnenborg, Rob Verhoeven, Murat Firat, Hanneke van Laarhoven, and Laura Genga. Towards evidence-based analysis of palliative treatments for stomach and esophageal cancer patients: a process mining approach. In *International Conference on Process Mining*, pages 136–143. IEEE, 2021.
14. Gregor Stiglic, Primoz Kocbek, Nino Fijacko, Marinka Zitnik, Katrien Verbert, and Leona Cilar. Interpretability of machine learning-based prediction models in healthcare. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 10(5):e1379, 2020.
15. Héctor G van den Boorn et al. Source: a registry-based prediction model for overall survival in patients with metastatic oesophageal or gastric cancer. *Cancers*, 11(2):187, 2019.
16. Héctor G van den Boorn et al. Source: prediction models for overall survival in patients with metastatic and potentially curable esophageal and gastric cancer. *Journal of the national comprehensive cancer network*, 19(4):403–410, 2021.
17. Maikel L Van Eck, Xixi Lu, Sander JJ Leemans, and Wil MP Van Der Aalst. Pm: a process mining project methodology. In *International conference on advanced information systems engineering*, pages 297–313. Springer, 2015.
18. Mozhgan Vazifehdoostirani, Mohsen Abbaspour Onari, Isel Grau, Laura Genga, and Remco Dijkman. Uncovering the hidden significance of activities location in predictive process monitoring. In *International Conference on Process Mining*, pages 191–203. Springer, 2023.
19. Mozhgan Vazifehdoostirani, Laura Genga, Xixi Lu, Rob Verhoeven, Hanneke van Laarhoven, and Remco Dijkman. Interactive multi-interest process pattern discovery. In *International Conference on Business Process Management*, pages 303–319. Springer, 2023.
20. Andrew J Vickers. Prediction models in cancer care. *CA: a cancer journal for clinicians*, 61(5):315–326, 2011.