# Towards Evidence-Based Analysis of Palliative Treatments for Stomach and Esophageal Cancer Patients: a Process Mining Approach

Pam Pijnenborg*, Rob Verhoeven *‡, Murat Firat †, Hanneke van Laarhoven‡ and Laura Genga§
*Dept. of Research & Development, Netherlands Comprehensive Cancer Organisation, Utrecht, The Netherlands
Email: p.pijnenborg@iknl.nl, r.verhoeven@iknl.nl
†Carou Institute, Open University, Heerlen, the Netherlands
Email: murat.firat@ou.nl
‡Dept. of Medical Oncology, Amsterdam University Medical Centers, Amsterdam, the Netherlands.
Email: h.vanlaarhoven@amsterdamumc.nl
§ Eindhoven University of Technology, Eindhoven, The Netherlands
Email: l.genga@tue.nl

***Abstract*—Stomach and esophageal cancer are in the top ten most common cancers worldwide, both with high mortality rate. Approximately one-third of these patients have metastases at initial diagnosis and should receive personalized palliative care to improve their remaining life time. However, there is a lack of consensus about personalized palliative care options. This often leads to difficulties in determining the right treatment pathway for individual patients. This study investigates the application of process mining techniques on palliative care pathways for stomach and esophageal cancer to obtain an evidence-based understanding of which palliative treatments are commonly carried out in clinical practice and how they are associated with patients' survival time. Given the high variability of the treatment pathways, 'local models' are derived, rather than end-to-end process models, which are then validated with the aid of physicians. In addition, this study also investigates the use of predictive process monitoring techniques to predict patients' life expectancy. The results show the benefit of taking the process-flow into account in predicting the outcome of the palliative treatments.**

***Index Terms*—Local Process Mining; Predictive Process Monitoring; Healthcare Processes**

## I. Introduction

Stomach and esophageal cancer (combined as esophagogastric cancer, EGC) are in the top ten most common cancers worldwide, both with high mortality rate [1], [2]. At initial diagnosis, almost a third of the esophageal and stomach patients have metastases [3]. Patients with metastatic EGC can usually not be cured and receive palliative care to increase the quality of the remaining life time and possibly to extend it. The disease course of EGC is however heterogeneous. Patients can have various complaints due to either the primary tumor (e.g. dysphagia, obstruction of food passage through stomach) or metastases (e.g. pain, organ dysfunction), with treatment options from different medical disciplines. Personalized multidisciplinary palliative care to treat the individual needs of these patients is thus urgently needed. However, nowadays there is no clear consensus on guidelines for the prescription of palliative care treatments for patients with EGC. One of the reasons can be found in the limited knowledge available on the most effective treatments for these patients. For treatment of specific symptomatology, few randomized clinical trials (RCTs) are available. For treatment with systemic therapy RCTs are often based on strict inclusion criteria regarding, e.g., physical fitness, or comorbidities of the patient which, while increasesing the rigor of the study, hampers the generalizability of the results [4], [5], [6]. The lack of evidence-based guidelines makes selecting the best option for these patients challenging [3], [7], [8], which results in a strong variability in treatments provided by different hospitals and different doctors. This highlights the need for developing more evidence-based palliative care strategies, based not only on RCTs but taking into account also other kinds of data.

The goal of this research consists in investigating the use of historical data of palliative treatments for EGC patients to provide evidence-based insights on current treatments and their impacts on the patients' survival time. The study is conducted in cooperation with the Amsterdam University Medical Centers and with the Netherlands Comprehensive Cancer Organisation (IKNL), a Dutch research institute that focuses on reducing the impact of cancer, and which maintains the Netherlands Cancer Registry (NCR), one of the most extensive disease specific registries worldwide (https://iknl.nl/over-iknl). We focus on the following research questions (RQs):

- What are common practices in the palliative treatment of EGC patients in The Netherlands which impact patients' survival time?
- Given an event log of completed treatment pathways and the patient's survival time of each case, how to train a model that can accurately predict the life expectancy of a patient currently under treatment?

To address these questions, we investigate the application of *process mining* (PM). PM is a family of techniques aimed

at analyzing event logs tracking process executions to extract useful insights on the corresponding process. An event log is a set of traces (aka, cases), each containing activities executed for a given patient at particular timestamps (i.e., events). Several previous studies have demonstrated the potential of using process mining techniques to analyze healthcare processes, suggesting that these techniques can effectively support in answering the above-mentioned research questions [9], [10], [11], [12], [13]. In this study, we apply *local process model* (LPM) *mining* techniques to extract common practices of the palliative process; while we use *predictive process monitoring* (PPM) to predict patient's life expectancy. The obtained results prove the value of PM approaches to understand palliative EGC treatment. The extracted models were indeed able to represent well-known cancer treatments, validated by the doctors; furthermore, the prediction models achieved a reasonable level of performance, and highlighted the importance of the process control-flow as a predictive feature for these processes. It is worth noting that the access to the NCR allowed this study to take into account a much larger cancer population than the ones considered in previous studies on EGC.

The rest of this paper is organized as follows. Section II provides a brief overview on related work; Section III introduces the research methodology; Section IV presents the case study and the results; Section V discusses the limitations of this study, together with future research directions.

## II. Related work

Process mining aims at exploiting data stored in organizations' information systems to understand and enhance the corresponding processes [14]. During the last decades there have been an increasing number of applications of PM in the health-care domain, with promising results, as reported by different survey studies [9], [10]. An interesting trend emerging from such studies is that the majority of previous work focused on applying process discovery techniques in order to provide evidence-based insights on how a given healthcare process was actually carried out in reality. To address the high variability of these processes, mostly either abstraction techniques (e.g., the Heuristic Miner [15], or the Fuzzy Miner [16]) or clustering techniques (e.g., [17]) have been used. The first category aims at abstracting infrequent and noisy behaviors to derive process models showing only the core process behavior, while the latter aims at clustering together similar traces to generate a set of models from more homogeneous behaviors, in place of a single model. A different strategy to deal with variable processes consists in mining *local models*, i.e., extracting only a subset of process behaviors which are of interest according to some user-defined criteria. While a number of local model extraction techniques have been proposed in the last years [18], [19], [20], [21], the above-mentioned surveys [9], [10] show that this topic is still quite underrepresented within the healthcare domain, despite its potential.

While most previous applications of PM to healthcare focus on *off line* analysis, a recent, emerging trend is the increasing interest in applying predictive process monitoring techniques to provide *on line* support to healthcare processes. The core idea is to train a predictive model to generate predictions about what will happen for a running case [22]. This requires to encode the log traces in a feature vector suitable for training the model. A number of trace encoding approaches have been introduced in the last years (e.g., using aggregation functions on event attributes [23], or explicitly modeling the order of the events [24]), and applied in combination with different prediction models (e.g., classic machine learning techniques like decision trees [25], or deep learning approaches like LSTM [26]). Examples of applications within the healthcare domain are predicting violations of clinical guidelines [27], in-hospital mortality [28], or the fulfillment of a set of goals [29]. In this work, we will investigate the use of PPM techniques to predict patients' life expectancy. In addition to state-of-the-art trace encoding approaches, we will investigate the use of LPM as predictive features which, to the best of our knowledge, has not been used in previous work.

## III. Methodology

This research has been conducted following the principles of the PM$^2$ framework, a well-known process mining project methodology aimed at understanding and improving an organization's processes [30]. The framework involves six steps: a) *Planning*, where RQ are defined; b) *Extraction*, where event data are extracted; three *analysis and iteration steps*, i.e., c) *data processing*, d) *mining and analysis*, e) *evaluation*, where different PM techniques are applied and their results evaluated, typically in an iterative fashion; finally, the findings of the analysis are used for the f) *process improvement and support* phase. The formulation of the RQ has been discussed in Section I. As regards the data extraction, data from the IKNL registry have been used (more details in Section IV). The following subsections delve into the PM methods applied for each RQ. Note that phase *f* is out of scope for this project.

### A. Local Process Model Discovery

The first RQ focuses on uncovering common practices within palliative treatments. To this end, we apply a *local process model discovery* technique, of which the goal consists in extracting relevant process patterns that would likely go undetected in an end-to-end model in a highly variable context. In particular, we apply the technique proposed in [19], since, to the best of our knowledge, this is the only approach that allows to represent the most common control-flow constructs (e.g., sequence, concurrency, loops). There, LPMs are in the form of *process trees*, where each node is either a leaf node (i.e., an activity) or a non-leaf node (i.e., an operator, like sequence ($\rightarrow$), choice ($\times$), concurrency ($\wedge$)). Each non-leaf node has one or more children, which can be leaf or non-leaf nodes. The discovery of a LPM set is an iterative process involving four steps, namely the *generation* of a set of candidate LPMs, the *evaluation* according to a set of metrics, the *selection* of the candidates fulfilling user defined thresholds on the metrics and, finally, the *expansion* of each

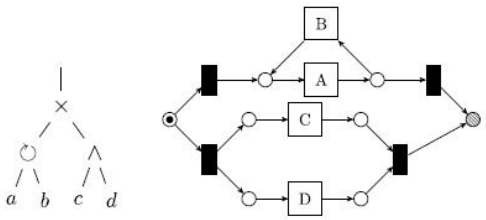


Fig. 1. A process tree and an equivalent petri net [19]

process tree with additional nodes. The process stops when no more candidates are generated or when a maximum iteration number is reached. Five metrics are used to assess the LPMs; the *support*, which represents the number of occurrences of the LPM; the *confidence*, that is the share of events of the activities in the LPM which fit the behavior described by the LPM; the *language fit*, that measures the ratio of behavior allowed by the LPM that is observed in the event log; the *determinism*, that relates to the degree to which future behavior can be determined; finally, the *coverage*, that measures the frequency of the activities described by the LPM in the log. The generated process trees are then transformed in *Petri nets*, a common process modeling language [31]. A Petri net is a bipartite graph consisting of *places* (used to represent states of process executions) and *transitions* (i.e., activities). Fig.1 shows a process tree and an equivalent Petri net.

In this study, we are interested in those LPMs which show a significant impact on the patients' survival time. To determine these high-impact LPMs, we make use of the *Cox Proportional Hazards Model*, a technique widely used in *survival analysis*, a field of statistics aimed at determining the overall survival time of a particular population under study [32]. The Cox Proportional Hazards Model is computed as follows [33] [34]:

$$h(t) = \beta_0(t) \exp\left(\sum_{i=1}^{n} b_i x_i\right) \quad (1)$$

where $h(t)$ represents the *hazard function*, that is the risk of dying at time *t*. The hazard function is determined by the product of the *baseline hazard function* $\beta_0$(t), that represents the hazard when all the covariates $x_i$ (i.e., factors that are expected to affect the survival time) are zero, and the exponential of the sum of the covariates with the corresponding coefficients $b_i$. The coefficients denote the impact of the covariates on the hazard function $h(t)$. The quantities $exp(b_i)$ are called *hazard ratios*; a value lower (higher) than one indicates that the covariate is associated with improved (decreased) survival. Note that survival times are not required to follow a specific statistical distribution, however, the proportion of the hazards for any two individuals should be constant over time [34]. In this study, the occurrence of LPMs are used as covariates to compute the Cox Proportional Hazards Model; hazard ratios are then used to quantify their impact on patients' survival. Note that the Cox model cannot be used to prove causal relations between the occurrence of a LPM and the observed patient's survival time. Nevertheless, it allows to identify LPMs statistically related with higher or lower survival time, thus providing useful insights on the treatment process. Note that data preprocessing is needed in order to remove possible *(multi) collinearities* before applying the Cox model [35]. Collinearity occurs when the independent variables in a regression model are strongly correlated and can lead to misleading results. In this study, we expect collinearity due mostly (though, not exclusively) to inclusion relations among LPMs (e.g., an LPM encompassed by a larger one). Our pre-processing takes into account both pairwise correlation and multicollinearity. To estimate pairwise correlation, we use the chi square test of independence [36]. When two non-independent LPMs are found, only one is kept (the one with highest support or, when the support is the same, the biggest one). As regards multicollinearity, we use the Variance Inflation Factor (VIF) [37], which is a well-known method in literature to mitigate colinearity. When collinearities are detected, the variable with the highest VIF score is removed from the dataset.

### B. Predictive Process Monitoring

The second RQ of this project focuses on predicting patients' life expectancy. In particular, we are interested in assessing to which extent the treatments a patient has been going through (i.e., the control-flow aspect of the treatment process) contribute to predict patients' life expectancy. Note that life expectancy in this study is treated as a categorical indicator, whose values represent temporal windows defined (and labeled) together with experts from IKNL (see Section IV-C). Therefore, outcome-oriented predictive process monitoring techniques are used. The standard PPM framework consists of the offline phase, where prediction models are trained on historical cases, and of the online phase, where these models are used to generate predictions on running cases. The offline phase consists of three steps.

*a) Create prefix buckets:* First, prefixes (i.e., trace subsequences of a predefined length) are generated from the event log and assigned to *buckets*, i.e., prefix groups. In this research, prefixes of equal prefix length are considered per bucket, a strategy often used in literature and that has shown good results in previous work [22]. We consider only prefix lengths with more than 100 patients, to ensure to have a reasonable number of samples supporting the results of the analysis.

*b) Sequence encoding techniques:* In this research, we tested the most commonly used sequence encoding techniques, which are briefly explained in the following by using the example trace $\sigma_1 = \langle consultation, ultrasound, ultrasound \rangle$, labeled as $c_{\sigma_1}$. The *boolean encoding* method encodes the activities with a 1 if the activity occurs in the trace, 0 otherwise. The *frequency-based* encoding method uses, instead, the frequency of each activity. Table I shows both these encoding for $\sigma_1$. These methods do not take the order of events into account. The *simple index* encoding [24], instead, considers the order by generating a feature for each activity position in the case. Table II shows the simple index encoding for $\sigma_1$. However, when the variability of the traces increases, the posi-

tion encoding may generate a sparse matrix, thus affecting the prediction performance. To address this issue, in this research we also investigate the use of LPMs as encoding strategy. They can provide a compact representation of core process behaviors, possibly involving also parallelisms and loops, thus allowing to encode relevant control-flow information in a smaller number of features than the simple index encoding. However, this compact representation comes at the expenses of possible information loss on specific event sequences. Every LPM is treated as a feature, with a binary encoding (i.e., 1 if the LPM occurrs for a patient, 0 otherwise). Table III shows the LPM-based encoding for $\sigma_1$, assuming that $LPM1 = ((consultation, ultrasound), \rightarrow)$, i.e., $ultrasound$ occurrs after $consultation$. Patients' attributes are encoded as well, using one-hot encoding for categorical attributes.

*c) Train prediction models:* In this stage, prediction models are trained on the prefix buckets. For this research, classic machine learning approaches have been considered. We plan to investigate the use of approaches based on Deep-learning in future work. We have selected Random Forest, XGBoost (highlighted as best prediction models in a recent benchmark study [22]), and decision trees. The latter technique was chosen because it is in general appreciated for being explainable for domain experts. Both a binary class and a multi-class setting has been tested to assess the impact of different discretization strategies on the prediction results. To optimize the prediction performance, both feature selection and hyperparameter optimization have been used. In particular, we tested two feature selection and extraction strategies, i.e. *wrapping methods* and *principal component analysis* (PCA) respectively. Wrapping methods select the combination of features leading to the highest performance of a selected classifier. While these methods can be computationally expensive, they perform in general better than other feature selection methods [38]. While wrapping methods return the optimal subset of features, PCA first selects principal components and uses these principal components to create new features [39]. We tested both methods aiming at obtaining new feature sets involving respectively 30, 50, and 75 percent of the original feature set. As regards hyperparameter optimization, we use Bayesian hyperparameter optimization, which uses a surrogate model that relies on earlier outcomes to select the most promising hyperparameters to evaluate in the objective function, and often proved to perfom better than grid and random search [40]. We selected the hyperparameters optimized per prediction model and their ranges (shown in Table IV) following what has been done in previous work [22], [41]. The hyperparameters are optimized for each feature set and predictive model. To obtain robust results, 3-fold cross validation is applied.

## IV. Case Study

### A. Event log description

The dataset used in this research is provided by the Netherlands Cancer Registry (NCR), maintained by the Netherlands Comprehensive Cancer Organisation (IKNL). The NCR dataset includes data of all newly diagnosed patients with cancer in the Netherlands since 1989, together with information related to their treatments. Patient data in the NCR is divided into multiple *episodes*, each representing a period of time until progression of the disease of the patient. Since life expectancy of most patients diagnosed with EGC is not very long, most of them receive only treatments in the first episode. Therefore, in this research the first episode is examined, ranging from 2015 until 2017. The data of the NCR fulfills the three requirements to be used in process mining analysis. The data of NCR contains a unique identifier (case ID) per disease per patient; the events can be derived from the data attribute that provides codes of treatments; the timestamp of the treatments is included (in days). Nevertheless, some data pre-processing is necessary, as explained in the following.

Some activities are logged in sequences in the event log, while according to domain experts they are executed in parallel. Therefore these activities are aggregated to high-level activities. A common example is for activities belonging to the same class of treatments which start within three days after the first treatment or before the stop of previously started treatments. In the data, this has been often observed for chemoterapy activities; for example, *Capecitabine* and *Oxaliplatine* have often been aggregated to the high-level activity *Capecitabine/Oxaliplatine*.

Furthermore, some activities have been relabeled, again in agreement with domain experts, to mitigate the variability. In particular, different *Radiotherapy* activities are relabeled to the activity *Radiotherapy*. Several low-frequency activities concerning radiotherapy on body parts where cancer has metastasized are relabeled to the single activity named *Rtmeta*. Other activities, including surgery where the esophagus or stomach's tumor is removed, are relabeled into the activity called *Chorg*. Lastly, activities concerning surgery on the metastases are relabeled into the activity termed *Chmeta*.

Finally, some cases were removed where there were logging errors (e.g., patients for which the survival was not known), as well as exceptional cases or outliers, like patients who received one or multiple treatment(s) abroad. Similarly, patients with too deteriorated health are removed from the dataset, as they are not fit enough to receive any treatment.

At the end of the pre-processing, the event log consisted of 2364 cases, 36 activities, and 6546 events. Analyzing process variants, we found that the event log contains 401 variants. Fig. 2 shows the frequency distribution of the top ten variants. It is worth noting that the first two variants alone represent almost a third of the cases. As regards the number of events per case, most cases (1200) consist of only one event. This was expected, since these patients usually do not live very long and, therefore, do not receive many treatments. Nevertheless, there is a large number of cases involving two or more treatments, which can be used to extract typical treatment combinations by means of LPM analysis.

TABLE I
BOOLEAN AND FREQUENCY ENCODING

| Trace | consultation | ultrasound | label |
|---|---|---|---|
| $\sigma_1$ | 1 | 1 | $c_{\sigma_1}$ |
| $\sigma_1$ | 1 | 2 | $c_{\sigma_1}$ |

TABLE II
SIMPLE INDEX ENCODING

| Trace | $e_1$ | $e_2$ | label |
|---|---|---|---|
| $\sigma_1$ | consultation | ultrasound | $c_{\sigma_1}$ |

TABLE III
LPM ENCODING

| Trace | LPM1 | label |
|---|---|---|
| $\sigma_1$ | 1 | $c_{\sigma_1}$ |

TABLE IV
PREDICTION MODEL, HYPERPARAMETER AND RANGE

| Prediction model | Hyperparameter | Range |
|---|---|---|
| Random Forest | Max features | [0, #features] |
| XGBoost | Learning rate | [0,1] |
| XGBoost | Subsample | [0.5,1] |
| XGBoost | Max tree depth | [4,30] |
| XGBoost | Colsample by tree | [0.5,1] |
| XGBoost | Min child weight | [1,6] |
| Decision Trees | Minimum samples split | [2,50] |
| Decision Trees | Minimum samples leaf | [2,50] |

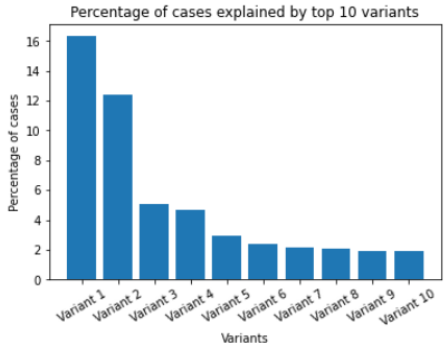


Fig. 2. Frequency of the ten most common process variants

### B. LPM mining results

*a) Settings:* We used the LPM miner plugin available in ProM 6.9 (https://www.promtools.org/doku.php). The user has to set three parameters to prune the LPM candidate set, i.e., the support, the event gap and the determinism level. The other metrics introduced in Section III are used to rank the obtained models. Note that there are no general guidelines; the parameters have to be tuned with respect to the problem at hand. In this respect, we selected a value of 10 for the support, following the domain experts' recommendations. The event gap determines how many events are allowed to occur in between the events corresponding to the LPM. Given the short length of most traces, we set the event gap to zero. We tested determinism values within a range between the default value of 0.5 until 0.8, with a step of 0.1, discussing the results with the domain experts. The LPMs generated with determinism 0.7 were chosen, since they offered a good trade-off between interpretability and generalization of the LPMs. Note that we did not use the inclusive choice operators when generating the LPMs, since it led to over-generalizing LPMs with little interpretability. From these settings, we obtained in total 53 LPMs, from which we removed the ones showing collinearity (see Section III-A); then, we applied the Cox hazard analysis on the remaining ones.

*b) Results:* The Cox Proportional Hazards Model's coefficients show that 17 LPMs exhibit a statistically significant impact on patients' overall survival time. More precisely, three LPMs show a negative impact, with an average hazard ratio of 2.02, while the rest show a positive impact,with an average hazard ratio of 0.47. We discussed the impact LPMs with oncologists from Amsterdam University Medical center. Most positive impact LPMs contain chemotherapy treatments. This is in line with doctors' expectations, according to which patients with esophageal and stomach cancer could be treated with chemotherapy to extend the expected overall survival if the patient is fit enough and willing to. As an example, Fig. 3 and Fig. 4 show two LPMs acknowledged by the physicians as a common combination of chemotherapy treatments. In particular, the first LPM contains a loop involving the chemotherapy activities *Capecitabine* and *Oxaliplatine*, which are the first-line chemotherapy treatments for palliative care for stomach and esophageal cancer patients according to medical protocols. The second one shows two treatments that are given in sequence, i.e., *Chorg* followed by *Chmeta*, which are invasive treatments and will only be given to patients who are fit enough and have specific disease characteristics; based on that, a positive association with survival is expected. Another interesting case regards the treatment called *Trastuzumab* (Fig. 5). Only patients that have an overexpression of HER2 in the tumor cells can be treated with *Trastuzumab* in addition to chemoterapy, and this is expected to have a positive impact on the survival as well. The presence of LPMs showing this combination and relating it to higher survival expectations confirms the medical knowledge.

Only a few LPMs with negative impact have been found and they were more tricky to interpret. Two representative ones are reported in Fig. 6 and Fig. 7. It is worth noting that these LPMs contain radiotherapy on metastases (*Rtmeta*). *Rtmeta* aims at reducing complaints and does not aim at enhancing overall survival, and it is often given to patients with worse health conditions. Therefore, the fact that these treatments are associated with worse survival expectancy is also in line with medical knowledge. However, this treatment seems to often be combined with chemotherapy treatments, which seem to be contradicting with the insights gained from the positive impact LPMs. For instance, *Capecitabine* and *Oxaliplatine* in parallel has a good impact, but in conjunction with *Rtmeta* the impact becomes negative. A possible explanation suggested by doctors is that possibly *Capecitabine* and *Oxaliplatine* are only

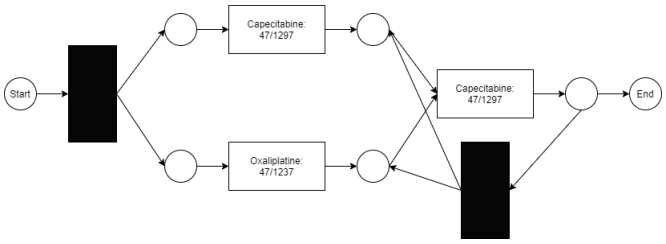


Fig. 3. Good impact: LPM14

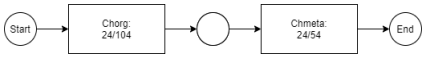


Fig. 4. Good impact: LPM23

given a short duration to the patients, so these activities did not have much impact. This suggest that in further research the duration of activities should also be considered.

Wrapping up, the applied technique has been able to extract from the data valuable insights on the palliative treatment process, returning common treatments, together with an assessment of their impact on patients' survival, which are in line with the medical knowledge. Nevertheless, further research is needed to address some limitations, like the lack of temporal information in the models, and to get a deeper understanding of the treatment impact. In this respect, an interesting direction consists in differentiating the impact of the LPMs with respect to patients' characteristics. While only patients with minimum fit conditions to get at least some treatment have been selected, some treatments have stricter minimum fit conditions to be given to the patients, making it interesting to further investigate this aspect. Nevertheless, these results represent the first steps towards the building of a shared, evidence-based knowledge on palliative EGC treatment that is necessary in order to deliver the best care option to each patient, as discussed in Section I.

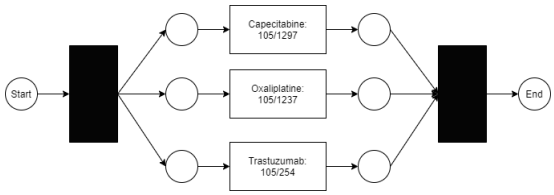


Fig. 5. Good impact: LPM8

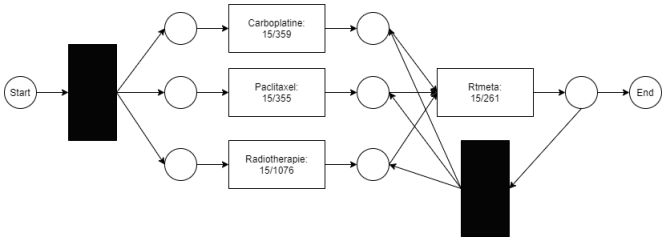


Fig. 6. Bad impact: LPM36

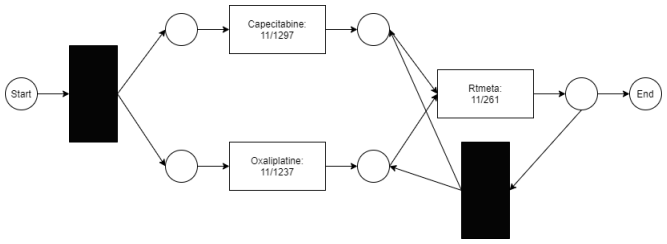


Fig. 7. Bad impact: LPM51

TABLE V
FEATURE SETS WITH NUMBER OF FEATURES

| Feature set | Number of features |
|---|---|
| Simple index encoding | 62 |
| Binary encoding | 45 |
| Frequency encoding | 45 |
| Impact LPMs | 17 |
| LPMs | 53 |
| Patient characteristics | 17 |

### C. Predictive Process Monitoring

To generate the life expectancy label for each patient, the corresponding survival time (recorded as number of days) has been divided into classes with equal frequency. Both two and three class setting have been tested. In the three class cases, the labels are 'Low', 'Middle', and 'High' corresponding to (30, 162], (162, 343], and (343, 1850] days of survival. For two classes, we created 'Low' and 'High' labels, corresponding to (30, 243] and (243, 1850] days. Class boundaries have been discussed with experts from IKNL. As discussed in Section III-B, different encoding strategies are tested; Table V shows the number of features obtained for each of them. Note that models are trained only for prefixes of length 1 and 2, since too few data were available for longer prefixes. Also, since LPMs consist of at least two activities, only for prefix length 2 the LPM features can be used. We consider both the set of all LPMs and the set involving only those LPMs that have an impact on patients' survival (referred to as "Impact LPMs").

*a) Results:* For the sake of space, here we only delve into the best results, obtained for two classes and prefixes of length 2. Table VI shows the highest F1 and accuracy score per combination of feature sets, classifier and feature selection method (FS), with percentage of features used. In addition to test each encoding strategy with/without considering the patients' characteristics, we were also interested in assessing to which extent the use of LPMs as predictive features was able to improve the classification performance w.r.t. the other encoding strategies. To this end, we tested also combinations of each LPM feature set with every feature set generated by the other encoding methods. The crosses in Table VI show the tested combinations. All the tested configurations perform similarly in terms of F1 score, but differences in terms of accuracy are more evident. In general, accuracy values are below F1 score, which suggests that these models are better in predicting the true positive (i.e., the "High" class). The use of patients' characteristics already allowed to obtain good

TABLE VI
BEST PERFORMING CLASSIFIERS PER FEATURE SET

| Patient char | Index | Frequency | Binary | LPMs | Impact LPMs | FS | Classifier | Percentage | F1 | accuracy |
|---|---|---|---|---|---|---|---|---|---|---|
| X | | | | | | PCA | XGB | 75 | 0.82 | 0.71 |
| X | X | | | | | SFS | XGB | 50 | 0.84 | 0.73 |
| X | X | | | X | | SFS | XGB | 75 | 0.84 | 0.76 |
| X | X | | | | X | None | XGB | 100 | **0.87** | **0.80** |
| X | | X | | | | SFS | XGB | 30 | 0.85 | 0.76 |
| X | | X | | X | | None | XGB | 100 | 0.82 | 0.73 |
| X | | X | | | X | None | XGB | 100 | 0.85 | 0.78 |
| X | | | X | | | None | XGB | 100 | 0.83 | 0.76 |
| X | | | X | X | | None | XGB | 100 | 0.82 | 0.73 |
| X | | | X | | X | None | XGB | 100 | 0.85 | 0.78 |
| | X | | | X | | SFS | XGB | 30 | 0.80 | 0.69 |
| | X | | | | X | SFS | XGB | 30 | 0.84 | 0.73 |
| | | | X | X | | SFS | DT | 30 | 0.82 | 0.71 |
| | | | X | | X | SFS | DT | 30 | 0.82 | 0.71 |
| | | X | | X | | SFS | DT | 30 | 0.82 | 0.71 |
| | | X | | | X | SFS | DT | 30 | 0.82 | 0.71 |
| | | | | | X | SFS | RF | 75 | 0.86 | 0.78 |
| | | | | X | | SFS | XGB | 30 | 0.78 | 0.64 |
| X | | | | | X | SFS | RF | 75 | 0.86 | 0.78 |
| X | | | | X | | None | RF | 100 | 0.84 | 0.73 |
| | X | | | | | SFS | XGB | 30 | 0.84 | 0.73 |
| | | | X | | | None | DT | 100 | 0.82 | 0.71 |
| | | X | | | | None | DT | 100 | 0.82 | 0.71 |

F1 results, although with poor accuracy. The combination of patients' characteristics with both index and Impact LPMs encoding obtained the best performance, highlighted in bold, improving significantly the accuracy with respect to the baseline. Overall, the table shows that adding impact LPMs to the other encoding strategies enhances the prediction performance. Interestingly, the impact LPMs is the feature set that obtained the second-highest accuracy and F1 scores, with a feature space significantly reduced w.r.t. the other encoding strategies (see Table V).

As regards the other testing settings, in general the binary classification obtained better results than the three class problem, where a maximum F1 score of 0.57 was obtained. This was expected, since binary classification is clearly an easier problem than the multi-class one. Furthermore, the quality of the prediction results increases when moving from prefixes of length 1 to prefixes of length 2. This was also expected, since in the latter case more information is available. Nevertheless, classifiers on length 1 prefixes still managed to achieve a F1 score of 0.84, that is comparable to the best one obtained in the length 2 prefixes. The main difference is in the accuracy, whose best value for length 1 prefixes is 0.69, against the 0.8 achieved for length 2 prefixes. Nevertheless, these results show that already after one treatment it is possible to obtain somewhat reliable predictions. A trade-off should be chosen between the earliness and the reliability of the prediction.

These results show that the built prediction models can predict, with reasonably good performance, life expectancy of a patient at multiple moments in the treatment process. It also shows the value of taking the process aspects into account when building the prediction. Nevertheless, additional research is needed to further enhance the performance of the classification models before they can be employed to provide operational support for these processes. We highlight some promising research directions in the following section.

## V. CONCLUSION AND FUTURE WORK

This research investigated the application of PM techniques to a) determine common treatment practices of palliative care for patients with EGC, and b) built predictive models for patients' life expectancy. We applied LPM mining techniques to extract common treatment practices, then using using the Cox Proportional Hazards Model to determine the impact of the extracted LPMs on patient's survival time. To predict patients' life expectancy, predictive process monitoring techniques have been employed, investigating the use of LPMs as prediction features. The obtained results highlighted the importance of the control-flow dimension for the prediction, and showed the potential of the usage of LPMs, which allowed to obtain good performance while significantly reducing the feature space.

The obtained results are promising and show the potential of PM techniques to understand and, ultimately, to improve the delivery of palliative treatments. Nevertheless, some limitations were also highlighted during the research. First, this research focused on the first line of treatment (treatment until first progression of disease); future studies should consider also patients going through multiple lines of treatments. Possible relations between patients' health conditions and chosen treatment, as well as their impact on the expected survival, should be investigated as well, as mentioned in Section IV. Furthermore, we plan to investigate the impact of more complex event attributes, like textual attributes (e.g., oncologists' notes about patients' complaints), or temporal properties (e.g., the duration of the treatments). Other process indicators should also be considered for the prediction, for example indicators related to patients' quality of life. Finally,

we plan to investigate the use of Deep Learning techniques, like LSTM classifiers.

## REFERENCES


[1] W. Fang, H. Igaki, Y. Tachimori, H. Sato, H. Daiko, and H. Kato, "Three-field lymph node dissection for esophageal cancer in elderly patients over 70 years of age," *The Annals of thoracic surgery*, vol. 72, no. 3, pp. 867–871, 2001.

[2] D. Ronald, "The age of cancer," *Nature*, vol. 408, no. 6809, pp. 248–254, 2000.

[3] W. P. Dijksterhuis, R. H. Verhoeven, M. Slingerland, N. Haj Mohammad, J. de Vos-Geelen, L. V. Beerepoot, T. van Voorthuizen, G.-J. Creemers, M. G. van Oijen, and H. W. van Laarhoven, "Heterogeneity of first-line palliative systemic treatment in synchronous metastatic esophagogastric cancer patients: A real-world evidence study," *International journal of cancer*, vol. 146, no. 7, pp. 1889–1901, 2020.

[4] A. Lai, S. Lipka, A. Kumar, S. Sethi, D. Bromberg, N. Li, H. Shen, L. Stefaniwsky, and P. Brady, "Role of esophageal metal stents placement and combination therapy in inoperable esophageal carcinoma: A systematic review and meta-analysis," *Digestive diseases and sciences*, vol. 63, no. 4, pp. 1025–1034, 2018.

[5] F. Lordick, C. Mariette, K. Haustermans, R. Obermannová, and D. Arnold, "Oesophageal cancer: Esmo clinical practice guidelines for diagnosis, treatment and follow-up," *Annals of Oncology*, vol. 27, pp. 50–57, 2016.

[6] E. Smyth, M. Verheij, W. Allum, D. Cunningham, A. Cervantes, and D. Arnold, "Gastric cancer: Esmo clinical practice guidelines for diagnosis, treatment and follow-up," *Annals of oncology*, vol. 27, pp. v38–v49, 2016.

[7] D. Cunningham, N. Starling, S. Rao, T. Iveson, M. Nicolson, F. Coxon, G. Middleton, F. Daniel, J. Oates, and A. R. Norman, "Capecitabine and oxaliplatin for advanced esophagogastric cancer," *New England Journal of Medicine*, vol. 358, no. 1, pp. 36–46, 2008.

[8] E. Ter Veer, N. Haj Mohammad, G. van Valkenhoef, L. L. Ngai, R. M. Mali, M. C. Anderegg, M. G. van Oijen, and H. W. van Laarhoven, "The efficacy and safety of first-line chemotherapy in advanced esophagogastric cancer: a network meta-analysis," *Journal of the National Cancer Institute*, vol. 108, no. 10, p. djw166, 2016.

[9] E. Rojas, J. Munoz-Gama, M. Sepúlveda, and D. Capurro, "Process mining in healthcare: A literature review," *Journal of biomedical informatics*, vol. 61, pp. 224–236, 2016.

[10] T. G. Erdogan and A. Tarhan, "Systematic mapping of process mining studies in healthcare," *IEEE Access*, vol. 6, pp. 24 543–24 567, 2018.

[11] R. S. Mans, W. M. P. van der Aalst, and R. J. Vanwersch, *Process mining in healthcare: evaluating and exploiting operational healthcare processes*. Springer, 2015.

[12] E. Rojas, M. Sepúlveda, J. Munoz-Gama, D. Capurro, V. Traver, and C. Fernandez-Llatas, "Question-driven methodology for analyzing emergency room processes using process mining," *Applied Sciences*, vol. 7, no. 3, p. 302, 2017.

[13] A. P. Kurniati, O. Johnson, D. Hogg, and G. Hall, "Process mining in oncology: A literature review," in *2016 6th International Conference on Information Communication and Management (ICICM)*. IEEE, 2016, pp. 291–297.

[14] W. M. P. van der Aalst, "Data science in action," in *Process mining*. Springer, 2016, pp. 3–23.

[15] A. J. M. M. Weijters, W. M. P. van der Aalst, and A. K. Alves de Medeiros, "Process mining with the heuristics miner-algorithm," *Technische Universiteit Eindhoven, Tech. Rep. WP*, vol. 166, pp. 1–34, 2006.

[16] C. W. Günther and W. M. P. van der Aalst, "Fuzzy mining–adaptive process simplification based on multi-perspective metrics," in *International conference on business process management*. Springer, 2007, pp. 328–343.

[17] R. J. C. Bose and W. M. P. van der Aalst, "Trace clustering based on conserved patterns: Towards achieving better process models," in *International Conference on Business Process Management*. Springer, 2009, pp. 170–181.

[18] R. P. J. C. Bose and W. M. P. van der Aalst, "Abstractions in process mining: A taxonomy of patterns," in *International Conference on Business Process Management*. Springer, 2009, pp. 159–175.

[19] N. Tax, N. Sidorova, R. Haakma, and W. M. P. van der Aalst, "Mining local process models," *Journal of Innovation in Digital Ecosystems*, vol. 3, no. 2, pp. 183–196, 2016.

[20] L. Genga, M. Alizadeh, D. Potena, C. Diamantini, and N. Zannone, "Discovering anomalous frequent patterns from partially ordered event logs," *Journal of Intelligent Information Systems*, vol. 51, no. 2, pp. 257–300, 2018.

[21] M. Acheli, D. Grigori, and M. Weidlich, "Efficient discovery of compact maximal behavioral patterns from event logs," in *International Conference on Advanced Information Systems Engineering*. Springer, 2019, pp. 579–594.

[22] I. Teinemaa, M. Dumas, M. L. Rosa, and F. M. Maggi, "Outcome-oriented predictive process monitoring: Review and benchmark," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 13, no. 2, pp. 1–57, 2019.

[23] F. M. Maggi, C. Di Francescomarino, M. Dumas, and C. Ghidini, "Predictive monitoring of business processes," in *International conference on advanced information systems engineering*. Springer, 2014, pp. 457–472.

[24] A. Leontjeva, R. Conforti, C. Di Francescomarino, M. Dumas, and F. M. Maggi, "Complex symbolic sequence encodings for predictive monitoring of business processes," in *International Conference on Business Process Management*. Springer, 2016, pp. 297–313.

[25] F. Folino, M. Guarascio, and L. Pontieri, "A prediction framework for proactively monitoring aggregate process-performance indicators," in *2015 IEEE 19th International Enterprise Distributed Object Computing Conference*. IEEE, 2015, pp. 128–133.

[26] N. Tax, I. Verenich, M. La Rosa, and M. Dumas, "Predictive business process monitoring with LSTM neural networks," in *International Conference on Advanced Information Systems Engineering*. Springer, 2017, pp. 477–492.

[27] C. Di Francescomarino, M. Dumas, F. M. Maggi, and I. Teinemaa, "Clustering-based predictive process monitoring," *IEEE transactions on services computing*, 2016.

[28] J. Theis, W. Galanter, A. Boyd, and H. Darabi, "Improving the in-hospital mortality prediction of diabetes icu patients using a process mining/deep learning architecture," *IEEE Journal of Biomedical and Health Informatics*, 2021.

[29] C. Di Francescomarino, C. Ghidini, F. M. Maggi, and F. Milani, "Predictive process monitoring methods: Which one suits me best?" in *International Conference on Business Process Management*. Springer, 2018, pp. 462–479.

[30] M. L. van Eck, X. Lu, S. J. Leemans, and W. M. P. van der Aalst, "PM$^2$: a process mining project methodology," in *International Conference on Advanced Information Systems Engineering*. Springer, 2015, pp. 297–313.

[31] W. Reisig, *Petri nets: an introduction*. Springer Science & Business Media, 2012, vol. 4.

[32] R. G. Miller Jr, *Survival analysis*. John Wiley & Sons, 2011, vol. 66.

[33] D. R. Cox, "Regression models and life-tables," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 34, no. 2, pp. 187–202, 1972.

[34] M. J. Bradburn, T. G. Clark, S. B. Love, and D. G. Altman, "Survival analysis part ii: multivariate data analysis–an introduction to concepts and methods," *British journal of cancer*, vol. 89, no. 3, pp. 431–436, 2003.

[35] K. P. Vatcheva, M. Lee, J. B. McCormick, and M. H. Rahbar, "Multicollinearity in regression analyses conducted in epidemiologic studies," *Epidemiology (Sunnyvale, Calif.)*, vol. 6, no. 2, 2016.

[36] T. M. Franke, T. Ho, and C. A. Christie, "The chi-square test: Often used and more often misinterpreted," *American Journal of Evaluation*, vol. 33, no. 3, pp. 448–458, 2012.

[37] E. R. Mansfield and B. P. Helms, "Detecting multicollinearity," *The American Statistician*, vol. 36, no. 3a, pp. 158–160, 1982.

[38] J. Li, K. Cheng, S. Wang, F. Morstatter, R. P. Trevino, J. Tang, and H. Liu, "Feature selection: A data perspective," *ACM Computing Surveys (CSUR)*, vol. 50, no. 6, pp. 1–45, 2017.

[39] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley interdisciplinary reviews: computational statistics*, vol. 2, no. 4, pp. 433–459, 2010.

[40] M. Binder, J. Moosbauer, J. Thomas, and B. Bischl, "Multi-objective hyperparameter tuning and feature selection using filter ensembles," *stat*, vol. 1050, p. 13, 2020.

[41] R. G. Mantovani, T. Horváth, R. Cerri, S. B. Junior, J. Vanschoren, and A. C. P. d. L. F. de Carvalho, "An empirical study on hyperparameter tuning of decision trees," *arXiv preprint arXiv:1812.02207*, 2018.