Predicting outpatient process flows to minimise the cost of handling returning patients: A case study

Marco Comuzzi[1], Jonghyeon Ko[1], and Suhwan Lee[1]

School of Management Engineering
Ulsan National Institute of Science and Technology (UNIST)
Ulsan, Republic of Korea
{mcomuzzi,whd1gus2,ghksdl6025}@unist.ac.kr

Abstract. We describe an application of process predictive monitoring at an outpatient clinic in a large hospital. A model is created to predict which patients will wrongly refer to the outpatient clinic, instead of directly to other departments, when returning to get treatment after an initial visit. Four variables are identified to minimise the cost of handling these patients: the cost of giving appropriate guidance to these returning patients, the cost of handling patients taking a non-compliant flow by wrongly referring to the outpatient clinic, and the false positive/negative rates of the predictive model adopted. The latter determine the situations in which patients have not received guidance when they should have had or have been guided even though not necessary, respectively. Using these variables, a cost model is built to identify which combinations of process intervention/redesign options and predictive models, when deployed together, are likely to minimise the cost overhead of handling these returning patients.

Keywords: Predictive monitoring, business process, case study, outpatient clinic.

1 Introduction

Predictive monitoring of business processes aims at developing techniques to predict aspects of interests of a process using data stored in so-called process event logs. The aspects predicted range from the next activity in a running case, to timestamps, that is, the instant of occurrence of future events in running cases, or process outcomes, such as satisfaction of service level objectives or constraints predicted over the order and occurrence of activities in running cases. Most frequently, predictive monitoring models are obtained by training machine learning models using data in event logs, appropriately pre-processed, as training and test sets. By empowering decision makers with insights about the possible occurrence of specific situations in the future, predictive monitoring supports proactive decision making, such as developing early warning systems or supporting resource reallocation and rescheduling.

A large number of techniques focused on different predictive monitoring tasks have been developed. These often focus on generating new features from event logs that may improve the accuracy of existing methods [6]. From an evaluation standpoint, the developed techniques are tested on a variety of publicly available event logs commonly used by the research community for benchmarking. Generally, the application of these techniques to specific contexts for the solution of real world problems remains limited. Commercial process mining tools do not normally incorporate advanced predictive monitoring into their functionality, and there are only very few examples of academic publications that have applied predictive monitoring in real world settings, particularly in the health care.

In this paper, we discuss an application of process predictive monitoring in a real world scenario, that is, to predict outpatient flows at a large hospital in South Korea. While this outpatient clinic strives to complete patient treatments in one single day, inevitably some patients, after an initial visit, are required to come back on a different day to receive another treatment. However, when returning, instead of referring directly to the department handling this additional treatment, many patients refer again to the reception desk of the outpatient clinic. This creates unnecessary load for the reception desk and, at the same time, it deteriorates patient satisfaction and level of service. The reception desk, in fact, often struggles to quickly redirect patients to the appropriate department from which they should receive treatment.

We build a simple model for minimising the overhead cost of handling patients wrongly returning to the outpatient clinic reception desk. The model considers, as cost-related variables, the cost of giving appropriate guidance to returning patients (in order for them to refer directly to the correct department when returning) and the cost of handling patients not referring directly to the treatment department. Additionally, we also devise two different classes of predictive models to predict whether patients will wrongly refer to the reception desk of the outpatient clinic. Models in the first class are trained using the event log as extracted from logging information of systems at the hospital. These models show high accuracy, but also a large rate of false positives. Being the returning patients a minority of cases, the second class of models is trained using oversampling of this minority class in the original event log. Models in this class achieve lower accuracy, but also lower false positive rates. Finally, considering the performance of the predictive models and different redesign options, which result in different relative values of the cost variables, the model can be used to identify the combinations of redesign options and predictive monitoring models that are likely to minimise the cost overhead of handling returning patients for the hospital.

The paper is organised as follow. Section 2 introduces the case study. Section 3 describes the methods and experimental settings used in our predictive monitoring analysis, while Section 4 presents the model of overhead costs of returning patients. The experimental results are discussed in Section 5. Related work is briefly outlined in Section 6 and conclusions are drawn in Section 7.

2 Case description

We consider the treatment process of outpatients at the Pusan National University Hospital (PNUH), which is one of the largest tertiary hospitals in Korea. The hospital hosts 42 medical departments and 341 rooms with 1351 beds. With a 25% increase of total outpatient visits in the last 5 years, quality of outpatient service has been issued as a critical agenda item at PNUH. The Korea National Patient Survey Program, firstly implemented in 2018, indicates that outpatients at PNUH experience low quality of clinical services, especially as far as the treatment process is concerned. As a consequence, PNUH has recently started a collaboration with academic partners to analyse and improve outpatient operations. The work presented in this paper is part of this initiative, and it relies on the analysis of an event log of outpatient visits recorded in a period from December 2017 to September 2018.

Table 1: Attributes of event log

Variable	*Definition*	*Data type*
caseID	case Combined patient id and visit number	Categorical
patientID	ID of a patient	Categorical
Time stamp	Complete timestamp of event	Date
Activity	Name of executed activity	Categorical
Resource	ID resource executing activity	Categorical
Resource_department	Department of the resource	Categorical
Clinic_department	Department providing treatment	Categorical
Disease_code	Type of diagnosed disease	Categorical
DoctorID	ID of doctor giving treatment	Categorical
Clinic room	Room number for consultation with a doctor	Categorical
Remaining time by next visit	Time remaining before the next visit	Continuous

Starting from logs of systems at the outpatient clinic and other departments, an event log has been firstly extracted with records of 145,638 outpatient visits at PNUH between December 2017 and September 2018 in 10 departments. As shown later, the department chosen for this case study are the 10 with the highest share of patients following a non-compliant flow when returning for a visit. The event log considers standard attributes, such as a case id, patient id, activity label, and timestamp, and other domain specific categorical and numerical attributes, as shown in Table 1.

As mentioned in the Introduction, the work presented in this paper is part of a larger initiative to identify and fix issues with outpatient operations. In particular, PNUH requires solutions to handle increasing number of patients, which cause low quality of services. The objective, in this case, is to improve the clinical processes under the constraint that the number and type of resources, e.g., nurses, doctors and admin staff, remains unchanged.

The first step, in our case, has been process discovery and, in particular, reaching consensus on a process model of outpatient visits with all stakeholders. Even though the outpatient clinic process at PNUH has slightly different flows for each department, activities in the process are recorded using common labels across departments. This has allowed us to integrate logs from 10 different departments to discover an overall common process. After a preliminary process discovery phase (using the process mining tool Disco), the standard process equally followed by all 10 departments has been identified in several meetings with stakeholders at PNUH, including nurses, IT staff, other administrative personnel, and doctors. Consensus was reached on the process shown in Fig. 1. The process starts with a patient referring to the reception desk of the outpatient clinic after having reserved an appointment online or by phone. During patient reception, administrative staff schedule an appointment with the department where a patient needs treatment, and instructs them on where and when to wait for such an appointment. Then, the patient goes to a waiting room in the specific department. Then, administration staff calls the patient for consulting with a doctor. During the consultation, the doctor may decide that the patient needs to revisit for an additional medical examination on a different day, if needed. After finishing the consultation, the patient pays for the visit and then may print a certificate of the treatment or get a prescription before returning home. For patients who have to visit again the hospital to get treatment, the patient should directly refer to the department providing the required treatment (the revisit part of the process is not depicted in Figure 1).

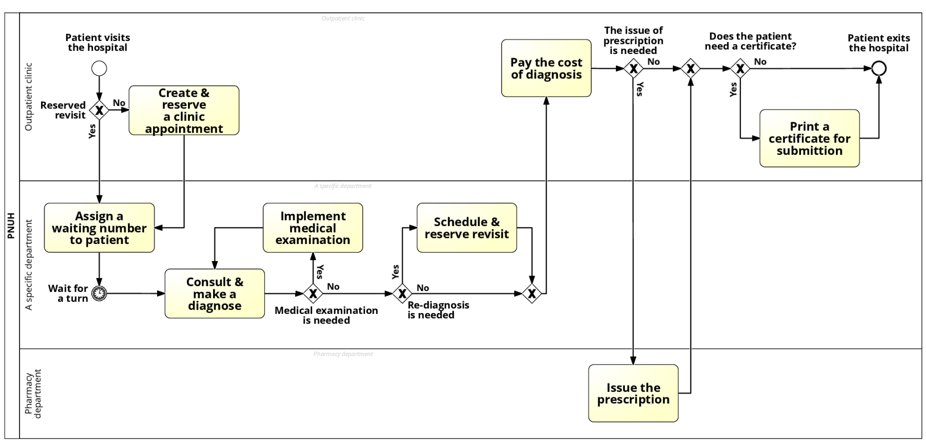

Fig. 1: BPMN 2.0 model for outpatient clinic process at PNUH hospital

In this paper, we investigate a particular problem of outpatient operations, i.e., non-compliant process cases in which patients, when returning to receive special treatment after a visit, wrongly refer to the reception at the outpatient clinic, rather than directly at the reception of the department where they should receive treatment (see Fig. 2, which models compliant and non-compliant process flows from a patient point of view). This results in disruption for the department giving treatment and longer waiting times for patients, since it is problematic for the outpatient clinic to quickly refer them to the correct department. A solution to this problem is to implement a predictive monitoring system to identify, at the moment a next visit is scheduled, and based on historic data, which patients are likely to wrongly refer to the outpatient clinic after the initial visit, i.e., following the non-compliant process flow. If an accurate model could be developed, then an appropriate guidance system could be put in place (e.g., sending reminders and/or instructing nurses/doctors to clarifying the correct referral procedure) to substantially reduce the likelihood of patients wrongly referring to the outpatient reception desk at the next visit.

To achieve this goal, the second step in the case study has been to quantify the number of patients following the non-compliant flow in the process. Fig. 2 shows that, across all departments, 17% of patients referred to the outpatient clinic reception desk rather than the specialist department when returning for a visit. Using information in the attribute *Remaining time by next visit* in the event log, we calculated that patients in this group spend on average 69 minutes more waiting to receive treatment compared to 83% of patients referring directly to the correct department when returning.

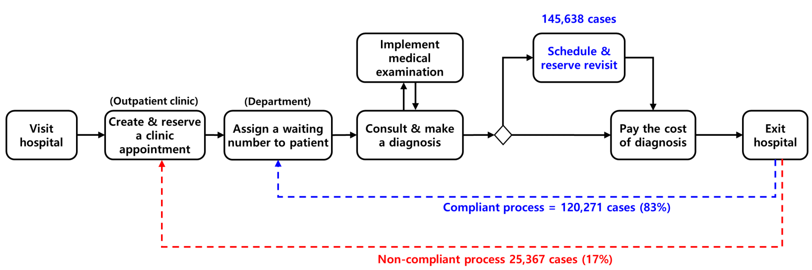

Fig. 2: Compliant and non-compliant process of returning outpatients

Table 2 shows the ratio of compliant and non-compliant cases for the 10 departments that we consider in this case study. In Section 5, results will be discussed for Department A and B, i.e., the two departments with highest number of non-compliant cases and, therefore, more critical from a performance analysis/customer satisfaction point of view for PNUH.

Note that, because of a non disclosure agreement with PNUH, the event log used in this paper cannot be made public and we are not allowed to reveal the actual names of the departments involved in our analysis.

Table 2: Non-compliant flows in analyzed departments

Department (Anonymized)	Proportion of non-compliant cases
Department A	32.5%
Department B	27.4%
Department C	19.9%
Department D	19.5%
Department E	18.9%
Department F	18.3%
Department G	16.1%
Department H	14.8%
Department I	14.0%
Department J	14.0%

3 Building predictive monitoring models

In order to create a predictive model for the problem that we have identified, we pre-processed the event log to obtain a feature-label dataset to feed into a classification algorithm. The predictive model considered in this paper adopts a simple feature vector using, for each case, the attributes of the activity *Reserve next visit*. After extensive test, among all the activities in the process, this is the one whose attributes better predict the behaviour of returning patients. The label, in this case, is binary, capturing whether the patient has followed the compliant process flow (referring to the correct department) or the non-compliant one (wrongly referring to the outpatient clinic reception). We leave the adoption of more *process-oriented* predictive models, such as predicting compliant/non-compliant returns by using next activity prediction [8] or as a process outcome [9], to future work.

Using these feature vector and label, we build predictive monitoring models with two classification techniques, i.e., decision tree (DT) and random forest (RF). Note that tree-based classifiers are known to be high performing techniques for binary classification problems and have also been shown to be effective for classification problems that use data in event logs [8]. We develop decision tree classifiers with parameters $max\_depth = 30$ and $min\_split = 2$ using the C4.5 algorithm (R package *rpart*), which involves tree pruning that reduces misclassification errors due to noise, and can be applied with both continuous and categorical attributes [2]. For random forest, we consider the implementation in the R package *randomForest* with parameters of $num\_tree = 1000$ and $mtry(split) = 3$, which is one of the most efficient implementations from the perspective of prediction accuracy and time cost [3]. In order to validate whether the developed models are trained without bias, the two R packages (*rpart*, *randomForest*) have their own built-in cross validation algorithms, named *x-error* and *out-of-bag error*, respectively. These cross validation algorithms, which are designed for reasonable approximation of test error, stabilise the error rate efficiently during tuning. Therefore, the dataset is divided into train set and test set in proportion of 75:25 and the train set is used for both fitting and validating the two models with the built-in cross validation algorithms.

Since the labels of the classifier input are imbalanced, we also consider oversampling techniques to decrease the imbalance rate. In many practical situations, in fact, classifiers tend to perform better when there is a balance between samples across the classes that have to predicted. This because, if one label is much more frequent than others, then the classifier may be able to generalise only the samples in the majority class. In this work, we adopt random oversampling of the minority class [1].

The performance of classifiers is evaluated using standard measures. Given the number of true positives $\#TP$, true negatives $\#TN$, false positives $\#FP$, and false negatives $\#FN$ obtained on the test set, we consider the ratio of true positives $TP = \frac{\#TP}{\#TP+\#FN}$, ratio of true negatives $TN = \frac{\#TN}{\#TN+\#FP}$, ratio of false positives $FP = \frac{\#FP}{\#TN+\#FP}$, ratio of false negatives $FN = \frac{\#FN}{\#TP+\#FN}$; and then the accuracy $A = \frac{\#TP+\#TN}{\#TP+\#TN+\#FP+\#FN}$ and F-1 score $F = \frac{2\#TP}{2\#TP+\#FP+\#FN}$.

4 An overhead cost model of handling returning patients

Before presenting and discussing the experimental results, in this section, we devise a simple optimisation model, which in the next section is used to identify the combination of process intervention options and predictive monitoring models that are likely minimise the overhead cost of handling returning patients at the wrong location.

This model is characterised by 4 variables:

- CG: this is the cost associated with giving appropriate guidance to a patient who has to return for treatment on a different day, in order for them to refer to the appropriate department, rather than the outpatient clinic. The magnitude of this cost depends on the way in which this guidance is implemented. Options are discussed later in this section;
- CR: this is the cost associated with handling one patient referring to the outpatient clinic reception desk, instead of directly returning to the department providing the treatment required. This cost is implicitly determined by the fact that patients returning at the wrong location create disruption at both the outpatient clinic and the referral department. In fact, these patients will be late for their appointment at the referral department, which creates obvious issues with appointment scheduling and execution;
- FP: this is the false positive rate of the predictive model;
- FN: this is the false negative rate of the predictive model.

False positives, in this model, are those patients that have been predicted not to return at the wrong location - and, therefore, who have not been given appropriate guidance - but that, actually, will return for treatment on a different day at the wrong location, i.e., referring to the outpatient clinic reception test. These patients will have to be handled by the reception desk. Therefore, for each false positive patient, the hospital sustains a cost equal to CR;

False negatives are patients that have been predicted to return for treatment at a later date at the wrong location, but that, in the end, did not make this mistake, and referred correctly to the department providing the treatment that they needed. From a practical standpoint, the hospital will have to give guidance to these patients, as instructed by the predictive model. The cost of giving this guidance, however, can be seen as an unnecessary cost for PNUH, since guidance was not necessary in the end. Therefore, for each false negative patient, the hospital sustains an opportunity cost that equal to CG.

The objective of the hospital is to minimise the following overhead cost function, $OVH(CG, CR, FN, FP)$, which represents the cost overhead of handling improperly returning patients at the outpatient clinic:

$$OVH = CR \cdot FP + CG \cdot FN$$

Note that the variables CG and CR are cost variables and, as such, depend on process intervention, i.e., on how guidance and handling of returning patients are implemented in the process, respectively. The variables FN and FP depend on the performance of the predictive model that is adopted for the implementation of the predictive monitoring system.

Different types of process interventions for both (i) the guidance given to patients predicted to take the non-compliant flow and (ii) the handling of patients actually taking the non-compliant flow, i.e., wrongly referring at the outpatient clinic reception when returning, can be considered.

Regarding the guidance given to non-compliant returning patients, the following options, classified in order of cost and expected effectiveness (from low to high) can be identified:

- Improved guidance from doctor of initial visit: if signalled, from the predictive monitoring system, that a patient is likely to take the non-compliant flow, doctors can be instructed in real time to take particular care in explaining the correct procedure when returning to patients;
- Phone call reminders: every day, the reception desk may call the patients expected to return on that day to remind them the correct flow;
- Social media: the correct flow can be reminded through automated messages on the social network services preferred by returning patients;
- Monitors at reception desk: monitors with clear instructions for patients returning on the day can be installed at the outpatient clinic to ensure that, even if entering the outpatient clinic, patients returning patients may read from the monitor to which department they should actually refer.

Note that these options are not mutually exclusive, but they can be combined. Obviously, the more the options implemented, the higher the cost CG.

The options available are more limited and expensive to implement for handling of patients wrongly referring at the outpatient clinic reception desk when returning. Specifically, it may be possible to integrate to a different degree the reception desk information system with the scheduling system in each department. The integration can range from simple notifications to each department of patients referring to the reception desk at the outpatient clinic, to full scale integration where personnel at the reception desk could re-schedule treatments at different departments on the same day for this type of patients.

5 Experimental results and discussion

This section first presents the experimental results of the predictive models. Then, we discuss a simulation of the cost model introduced in Section 4.

Table 3 and 4 show the performance of the predictive model developed for department A and B, respectively. Results are reported for the original event log (default) and for 4 different oversampling configuration, with the minority class increasing from the actual value in the default event log to 60%. The performance in terms of accuracy and F1-score is remarkable, even on the default imbalanced event logs. While DT and RF show similar performance on the original logs, accuracy and F1-score in the oversampled configurations tend to be higher for RF models.

Table 3: Result of classification in department A

Event	Decision Tree			Randomforest		
log	(TP, TN, FP, FN) rate	ACC	F1-score	(TP, TN, FP, FN) rate	Accuracy	F1-score
Default (68:32)	(0.84, 0.42, 0.58, 0.16)	0.757	0.798	(0.90, 0.33, 0.67, 0.10)	0.741	0.813
OS1 (60:40)	(0.64, 0.69, 0.31, 0.36)	0.860	0.818	(0.72, 0.64, 0.36, 0.29)	0.911	0.840
OS2 (55:45)	(0.60, 0.73, 0.27, 0.40)	0.867	0.806	(0.69, 0.68, 0.32, 0.31)	0.921	0.837
OS3 (50:50)	(0.57, 0.75, 0.25, 0.43)	0.868	0.759	(0.65, 0.73, 0.27, 0.35)	0.922	0.831
OS4 (45:55)	(0.57, 0.74, 0.26, 0.43)	0.872	0.759	(0.63, 0.75, 0.25, 0.37)	0.927	0.829
OS5 (40:60)	(0.56, 0.76, 0.24, 0.44)	0.876	0.750	(0.59, 0.78, 0.22, 0.41)	0.935	0.824

Table 4: Result of classification in department B

Event	Decision Tree			Randomforest		
log	(TP, TN, FP, FN) rate	ACC	F1-score	(TP, TN, FP, FN) rate	Accuracy	F1-score
Default (73:27)	(0.84, 0.63, 0.37, 0.16)	0.857	0.850	(0.85, 0.66, 0.34, 0.15)	0.867	0.861
OS1 (60:40)	(0.78, 0.67, 0.33, 0.22)	0.816	0.717	(0.77, 0.80, 0.20, 0.23)	0.809	0.759
OS2 (55:45)	(0.75, 0.70, 0.30, 0.25)	0.826	0.698	(0.77, 0.83, 0.17, 0.23)	0.823	0.749
OS3 (50:50)	(0.67, 0.73, 0.27, 0.33)	0.831	0.674	(0.76, 0.83, 0.17, 0.24)	0.838	0.732
OS4 (45:55)	(0.67, 0.74, 0.26, 0.33)	0.822	0.672	(0.75, 0.85, 0.15, 0.25)	0.845	0.719
OS5 (40:60)	(0.66, 0.76, 0.24, 0.34)	0.834	0.688	(0.74, 0.87, 0.13, 0.26)	0.854	0.698

Regarding the variables FP and FN of the cost model, it is worth noticing that oversampling of the minority class (i.e., wrongly returning patients) of the default event log clearly helps to reduce the false positive rate. For instance, in the case of department A (Table 3), the evenly oversampled configuration (50:50) more than halves the ratio of false positive when compared with the default log. At the same time, oversampling of the minority class also increases the false negative rate FN, even though to a lower degree than the decrease of FP.

Table 5: Overhead cost of patients wrong returns, department A (minimum values in bold)

Method	Event log	Rate		Cost : $OVH = CR \cdot FP + 0.5 \cdot FP$				
		FP	FN	$CR = 0$	$CR = 0.25$	$CR = 0.5$	$CR = 0.75$	$CR = 1$
Decision Tree	Default (68:32)	0.58	0.16	**0.079**	**0.223**	0.368	0.512	0.657
	OS (60:40)	0.31	0.36	0.180	0.257	0.334	0.411	0.489
	OS (55:45)	0.27	0.40	0.198	0.266	**0.334**	0.402	0.470
	OS (50:50)	0.25	0.43	0.217	0.278	0.340	0.401	0.463
	OS (45:55)	0.26	0.43	0.216	0.282	0.348	0.414	0.479
	OS (40:60)	0.24	0.44	0.221	0.281	0.340	**0.400**	**0.459**
Random forest	Default (68:32)	0.67	0.10	**0.050**	**0.218**	0.387	0.556	0.724
	OS (60:40)	0.36	0.29	0.143	0.233	0.323	0.413	0.504
	OS (55:45)	0.32	0.31	0.157	0.236	0.315	0.394	0.473
	OS (50:50)	0.27	0.35	0.175	0.242	**0.310**	0.377	0.444
	OS (45:55)	0.25	0.37	0.187	0.249	0.310	0.372	0.433
	OS (40:60)	0.22	0.41	0.205	0.259	0.313	**0.367**	**0.421**

Table 6: Overhead cost of patients wrong returns, department B (minimum values in bold)

Method	Event log	Rate		Cost : $OVH = CR \cdot FP + 0.5 \cdot FP$				
		FP	FN	$CR = 0$	$CR = 0.25$	$CR = 0.5$	$CR = 0.75$	$CR = 1$
Decision Tree	Default (73:27)	0.37	0.16	**0.079**	**0.170**	**0.262**	0.354	0.445
	OS (60:40)	0.33	0.22	0.110	0.192	0.275	0.358	0.440
	OS (55:45)	0.30	0.25	0.123	0.199	0.274	**0.349**	0.424
	OS (50:50)	0.27	0.33	0.163	0.230	0.296	0.362	0.429
	OS (45:55)	0.26	0.33	0.164	0.228	0.292	0.356	0.420
	OS (40:60)	0.24	0.34	0.172	0.233	0.293	0.354	**0.414**
Random forest	Default (73:27)	0.34	0.15	**0.073**	**0.158**	0.243	0.328	0.413
	OS (60:40)	0.20	0.23	0.115	0.165	0.214	0.264	0.313
	OS (55:45)	0.17	0.23	0.117	0.160	0.203	0.246	0.289
	OS (50:50)	0.17	0.24	0.122	0.164	0.205	0.247	0.289
	OS (45:55)	0.15	0.25	0.125	0.164	0.202	0.240	0.279
	OS (40:60)	0.13	0.26	0.132	0.165	**0.198**	**0.231**	**0.264**

Table 5 and 6 show the results of a simulation of the cost of handling patients returning to the wrong location for Department A and B, respectively. In particular, we consider the rate FP and FN from the predictive models presented above. As far as the costs CR and CG are concerned, we fix the value of CG (arbitrarily chosen to be 0.5) and we calculate OVH for different values of CF ($a = 0, 0.5, 1, 1.5, 2$).

We can notice that as the cost CR increases in respect of CG, the event log with oversampled minority class becomes the one associated with lower overhead costs for handling patients at the wrong location. In practical terms, this means that, as the cost CR associated with patients following the non-compliant flow increases in respect of the cost CG of the process intervention chosen for giving guidance to patients predicted to take the non-compliant flow, PNUH will be better off by adopting a predictive model that largely oversamples non-compliant cases.

The choices of process intervention options depends naturally on a number of other factors not considered by this model, such as available budget, target customer satisfaction, or the actual effectiveness of the patient guidance intervention (in this work, in fact, we implicitly assumed that any intervention in this regard will work 100% effectively, so that all patients receiving guidance will refer to the correct department). However, it is worth noticing that, given a specific configuration of process interventions, the work presented in this paper can be used to suggest a predictive model (i.e., default or oversampling the minority class) that helps to minimise the overhead overhead cost of the hospital.

6 Related work

Within the process mining literature, process predictive monitoring has emerged recently as a new set of techniques that aim at adapting traditional data mining and machine learning techniques to predict aspects of interest in a process [6]. Process data in event logs are seen as a knowledge base, which is used to train a machine learning model to predict at runtime some aspects of interest of currently executing process cases. There are 3 classes of aspects of interests that can be predicted using predictive process monitoring: (i) the activities that will occur in a running process case in the future, (ii) the time at which activities in running process cases will occur and (iii) outcomes of processes. Predicting next activities in running process cases enables the design of warning systems, in which stakeholders can be warned should a critical or risky activity be predicted to happen with high likelihood in the near future. Classification techniques are adopted to solve this problem [9].

Predicting times of next activities can also be used to design warning systems, i.e., to warn stakeholders that delays in process execution can be expected, and to support pro-active decision making by having reliable estimates of when processes may terminate. Regression techniques are used to solve this problem [9]. Finally, outcome-based predictive monitoring deals with predicting outcome of running case. There are different examples of outcomes that can be predicted, such as whether given performance or service level objectives will be eventually met, or constraints predicted on the order and occurrence of activities in a running process case. Since different outcomes are usually categorical variables, classification techniques are used to solve this problem [10].

Business processes in the health care have often been target of process mining initiatives in the literature. This because they are usually very challenging, being characterised by high variability and unpredictability while, at the same time, being sufficiently digitised to make available data to compile high quality event logs [7]. As far as predictive analytics is concerned, simulation emerges as a techniques which is often used [5,4]. However, it is generally adopted as a complementary method to process mining for analysing and improving offline the care pathway, rather than for improving the online reaction to unexpected situations, which is the more natural application of predictive monitoring.

7 Conclusions

This paper has presented a case study about the application of predictive monitoring techniques for improving the handling of returning patients at an outpatient clinic in a large hospital in Korea. This work is currently being improved in several ways. First, we are extending the cost model, by considering, for instance, that different guidance intervention may be effective to a different degree in changing patient behaviour. Then, we are experimenting with more accurate prediction models, by applying different process-oriented paradigms, such as next activity and process outcome prediction. Finally, we are currently working with PNUH to translate the results of the proposed model into actionable process redesign options.

References

1. A. Ali, S. M. Shamsuddin, and A. L. Ralescu. Classification with class imbalance problem: a review. *Int. J. Advance Soft Compu. Appl*, 7(3):176–204, 2015.
2. M. N. Anyanwu and S. G. Shiva. Comparative analysis of serial decision tree classification algorithms. *International Journal of Computer Science and Security*, 3(3):230–240, 2009.
3. C. Crisci, B. Ghattas, and G. Perera. A review of supervised machine learning algorithms and their applications to ecological data. *Ecological Modelling*, 240:113–122, 2012.
4. O. A. Johnson, T. B. Dhafari, A. Kurniati, F. Fox, and E. Rojas. The clearpath method for care pathway process mining and simulation. In *International Conference on Business Process Management*, pages 239–250. Springer, 2018.
5. R. Konrad, K. DeSotto, A. Grocela, P. McAuley, J. Wang, J. Lyons, and M. Bruin. Modeling the impact of changing patient flow processes in an emergency department: Insights from a computer simulation study. *Operations Research for Health Care*, 2(4):66–74, 2013.
6. A. E. Marquez-Chamorro, M. Resinas, and A. Ruiz-Corts. Predictive monitoring of business processes: a survey. *IEEE Transactions on Services Computing*, 2017.
7. E. Rojas, J. Munoz-Gama, M. Sepúlveda, and D. Capurro. Process mining in healthcare: A literature review. *Journal of biomedical informatics*, 61:224–236, 2016.
8. B. A. Tama and M. Comuzzi. An empirical comparison of classification techniques for next event prediction using business process event logs. *Expert Systems with Applications*, 129:233–245, 2019.
9. 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)*, 13(2):17, 2019.
10. I. Verenich, M. Dumas, M. La Rosa, F. Maggi, and I. Teinemaa. Survey and cross-benchmark comparison of remaining time prediction methods in business process monitoring. *arXiv preprint arXiv:1805.02896*, 2018.