Artificial Intelligence in the Operating Room: evaluating traditional classifiers to predict patient readmission

Rita Sacramento
rita4sacramento@gmail.com
Instituto Superior de Engenharia de Coimbra

Rui Silva
rsilva@b-simple.pt
BSimple, Porto, Portugal

Inês Domingues
inesdomingues@gmail.com
Medical Physics, Radiobiology and Radiation Protection Group, IPO Porto Research Centre (CI-IPOP)

Abstract

The readmission of patients who had surgery is very prevalent. The goal of this work is to develop machine learning models to predict the likelihood of this readmission. The models will be based on several characteristics such as pathology, surgical speciality, surgical intervention, among others. Given a group of clinical cases, collected from 3 hospitals, the above mentioned parameters are collected and used to train and test machine learning algorithms such as Logistic Regression (LR), Support Vector Machines (SVM), K-Nearest Neighbour (kNN) and Decision Trees (DT). Data imputation and data balance techniques were also used. Models were developed with pre-surgery data only and also with data from after the surgery. Decision Trees have shown the best performance, having an accuracy of 91% before surgical intervention and an accuracy of 82% after surgical intervention.

1 Introduction

The analysis of the number of readmissions is of uttermost importance since, in addition to the added expenses for the hospital and the implications for the patient, it is also a marker of quality of the service provided by the healthcare facility. In a study carried out in Portugal between 2000 and 2008, it was possible to conclude that of the 5 514 331 unplanned hospitalisations, 4.1 % corresponded to hospital readmissions and that in episodes of readmission, hospital mortality was higher than in the remaining episodes, with the mortality rate in readmission episodes being 9.5 % and in the remaining 5.6 % [8].

The developed work consists in the application of machine learning techniques to a database of 21 112 occurrences, acquired in three different hospitals. Models were developed for two phases, before surgery, using a total of 7 attributes, and after surgery, using 13 attributes. Data imputation and data balance were performed and 4 classifiers were tested, Logistic Regression (LR), Support Vector Machines (SVM), K-Nearest Neighbour (kNN) and Decision Trees (DT). Decision Trees classifier is the one that has the best performance in both phases, achieving an accuracy of 0.91 before surgery and 0.82 after surgery.

2 Commercially available software

Companies like Jvion, AI Brisbane and Safecare AI use artificial intelligence to develop hospital software for decreasing or predicting readmission of patients.

Jvion [1] identifies and predicts patients at risk and defines actions to be taken for each patient. To reach a decision on the likelihood that a patient will be readmitted within 30 days, the data that is taken into account is not only clinical data, but also external data such as whether they have access to food, pharmacy or car. According to Jvion, self learning Eigen Spheres are used, although the details on this technique are not very clear [2]. According to the website, in a recent test, Jvion software was able to correctly identify patients at high risk of readmission 96% of the time.

AI Brisbane [3], focuses on forecasting using machine learning algorithms. It makes a selection and division of patients into groups, of high and low risk of readmission. In addition to the readmission forecast, the reason for this classification is also explained.

SafeCare AI [4] has a more preventive action, using real-time decisions where the focus is on actions that may decrease the likelihood of a next readmission. Medical data is processed using AI software to provide clinical decision support by intelligence emulation using machine learning, deep learning and artificial neural networks.

Unlike the above mentioned software, the software to be developed by BSimple is focused on the episode of the operation, assessing the risk of a certain patient being readmitted, before and after medical intervention.

3 Methods

This section will detail the steps taken during the development. The pipeline includes four main phases: pre-processing, training, evaluation and forecasting, as shown in Figure 1.

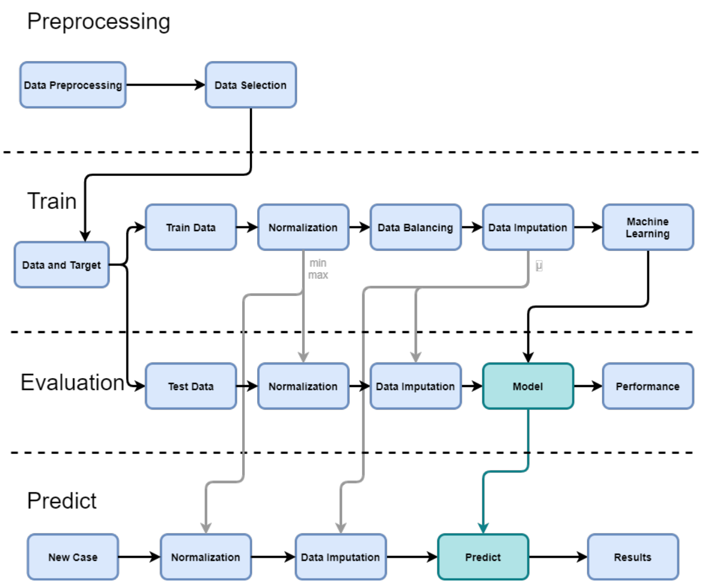

Figure 1: Methodology flowchart

Dataset: The used database has records from three Portuguese hospitals and contains 21,112 record. This database has 46 tables with different attributes. The attributes used in the model were empirically selected and depend if the instance corresponds to pre or post surgery.

Pre-processing: The database was first cleaned to remove all instances for which the surgery had been cancelled. To calculate which patients were readmitted, an interval of 30 days was considered.

Data selection: For the experiments before surgery, all cases with the variables: pathology, diagnoses, surgical speciality and surgical intervention with values equal to the case to be predicted were selected. These subgroups were then concatenated, removing the lines of repeated values. For the experiments after surgery, all cases with the variables: pathology, diagnoses, surgical speciality, surgical intervention, anaesthesia technique, and complications with values equal to the case to be predicted were selected. These subgroups were then concatenated, removing the lines of repeated values. This resulted in a set of 133 cases before surgery (79 non-readmissions and 54 readmissions) and 8 095 for after surgery (5 314 non-readmissions and 2 781 readmissions).

Data normalisation: The purpose of normalisation is to change the values of the data group used so that they all follow the same scale. The formula used is as follows: $z = \frac{x - min(x)}{max(x) - min(x)}$. After normalisation, the values now belong to the interval [0,1].

[1] https://jvion.com
[2] https://www.reddit.com/r/datascience/comments/8c2vnd/what_is_everyones_opinion_on_jvion_and_their/
[3] https://aibrisbane.com.au
[4] https://www.safecareai.com

Data balance: Data imbalance is characterised by a discrepancy in the number of examples per class of a dataset. This phenomenon is known to deteriorate the performance of classifiers, since they are less able to learn the characteristics of the less represented classes [3, 6]. In this way, before using the classifier, data was balanced using SMOTE (Synthetic Minority Over-sampling TEchnique) [1], applied only to training data.

Data imputation: Missing data has been found to have a considerable impact on the learning process of classifiers [7]. Since some fields were missing for some instances in our database, data imputation was performed. Some of the variables were filled in with the value zero, since this value is not in the list for the designation of any attribute. We thus assumed that null meant zero. For the variable "Hour", which only exists in the phase after the operation, the average calculated using only the training data (μ), was used to fill both the training and test data.

Classifiers: The current work deals with binary classification, being the two classes the readmission and no readmission of the patient. Four different classifiers were tested: LR, SVM, DT and kNN.

4 Evaluation

In order to test the performance of the classifiers, the dataset was divided into two sets, a train and a test set. A percentage of 30% was used for the test set, and the remaining was left for the train test. The evaluation methodologies used were the confusion matrix, accuracy, precision, f1-score and recall.

Precision, Recall and F1-Score before surgery are given in Table 1, while the same values for after surgery are summarised in Table 2. Accuracy results are shown in Table 3.

Table 1: Precision, Recall and F1-Score before surgery

Precision	LR	SVM	kNN	DT
Not readmitted	0.64	0.63	0.68	**1.00**
Readmitted	0.85	0.68	0.87	**0.96**
Recall	LR	SVM	kNN	DT
Not readmitted	0.92	0.76	0.92	**0.96**
Readmitted	0.46	0.54	0.54	**1.00**
F1-Score	LR	SVM	kNN	DT
Not readmitted	0.75	0.69	0.78	**0.98**
Readmitted	0.59	0.60	0.67	**0.98**

Table 2: Precision, Recall and F1-Score after surgery

Precision	LR	SVM	kNN	DT
Not readmitted	0.82	0.84	0.80	**0.87**
Readmitted	0.55	0.55	0.65	**0.75**
Recall	LR	SVM	kNN	DT
Not readmitted	0.68	0.76	0.76	**0.86**
Readmitted	0.72	0.73	0.65	**0.75**
F1-Score	LR	SVM	kNN	DT
Not readmitted	0.74	0.80	0.78	**0.86**
Readmitted	0.62	0.67	0.62	**0.75**

Table 3: Accuracy results

Accuracy	LR	SVM	kNN	DT
Before surgery	0.69	0.65	0.73	**0.98**
After surgery	0.69	0.75	0.72	**0.82**

The classifier with the best performance was Decision Trees, achieving an accuracy of 98% for the before surgery experiments and of 82% for after surgery. The worst classifier was SVM with an accuracy of only 65% for before surgery and of 75% after surgery.

Looking at the state of the art, Accuracy is between 69% and 72% in [4], while the Accuracy reported in [5] reaches values between 64% – 70%. It can be thus be concluded that our results exceeded the ones previously published.

For this scenario, it is more important to predict that a patient will be readmitted, even when he ends up not being readmitted. This prediction will allow for preventive measures to be undertaken. When analysing the confusion matrices (not shown due to space constrictions), it could be seen that, before surgery, only one case were miss-classified as "Not readmitted" when in fact they were readmitted within 30 days. For after surgery, 428 cases were miss-classified as "Not readmitted" when in fact they were readmitted.

Having selected the best model, Decision Trees, it is important to access its stability. In this way, a set of 30 runs was performed, each time with different randomly selected train and set sets (always in the proportion of 70%/30%). Average and standard deviation of the accuracy in each run is summarised in Table 4.

Table 4: Accuracy stability assessment

	Average	Standard deviation
Before surgery	0.91	0.03
After surgery	0.82	0.01

It can be seen that accuracy average values are very close to the ones previously stated, and standard deviation are low, assuring the models' stability.

5 Conclusions

The objective of this project was to develop a forecast model for the readmission of a patient before and after undergoing a surgical intervention. The existence of these models will have a high impact in the clinical practice. A patient predicted to be readmitted can be more carefully analysed by the healthcare staff and more tests and procedures can be performed before his release from the hospital in order to reduce the number of readmissions. Even after the release of the patient from the hospital, a closer monitoring of the recovery by phone calls or schedule appointments can be done to identify early possible problems.

Although the developed model is functional, there are improvements that could be made. The application of deep learning techniques [2] is a possibility. We note, however, that these type of models need to be carefully evaluated, being that simpler models are to be favoured in this context.

6 Acknowledgements

This work is partially financed by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência Tecnologia within project UIDP/00776/2020.

References

[1] NV Chawla, KW Bowyer, LO Hall, and WP Kegelmeyer. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16:321–357, 2002.

[2] I Domingues and J S Cardoso. Mass detection on mammogram images: a first assessment of deep learning techniques. In *19th Portuguese Conference on Pattern Recognition*, 2013.

[3] I Domingues, JP Amorim, PH Abreu, H Duarte, and J Santos. Evaluation of Oversampling Data Balancing Techniques in the Context of Ordinal Classification. In *IEEE International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2018.

[4] HJ Lai, PC Chan, HH Lin, YF Chen, CS Lin, and JC Hsu. A Web-Based Decision Support System for Predicting Readmission of Pneumonia Patients after Discharge. In *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 2305–2310, 2018.

[5] HJ Lai, TH Tan, CS Lin, YF Chen, and HH Lin. Designing a clinical decision support system to predict readmissions for patients admitted with all-cause conditions. *Journal of Ambient Intelligence and Humanized Computing*, 2020.

[6] F Marques, H Duarte, J Santos, I Domingues, JP Amorim, and PH Abreu. An iterative oversampling approach for ordinal classification. In *34th ACM/SIGAPP Symposium on Applied Computing*, pages 771–774, 2019.

[7] RC Pereira, MS Santos, PP Rodrigues, and PH Abreu. MNAR Imputation with Distributed Healthcare Data. In *EPIA Conference on Artificial Intelligence*, pages 184–195, 2019.

[8] B Sousa-Pinto, AR Gomes, A Oliveira, C Ivo, G Costa, J Ramos, J Silva, MC Carneiro, MJ Domingues, MJ Cunha, A Costa-Pereira, and A Freitas. Reinternamentos hospitalares em Portugal na última década. *Acta Medica Portuguesa*, 2013.