Tim Schneller
Inselweg 9
8630 Rüti ZH
tim94@hispeed.ch

Predicting Patient Satisfaction based on a collection of preoperative factors by applying machine learning algorithms on medical registry data

Conceptual Design Report

31 October 2023

Abstract

In orthopedic surgery, one of the most important outcome measures is patient satisfaction. There are many risk factors, that are beyond the control of the surgeon, which can influence the treatment, its' outcome and hence patient satisfaction. The purpose of this project is to provide a reliable predictive tool for surgeons to use in their daily clinical work, which predicts patient satisfaction two years after shoulder arthroplasty surgery, based on preoperative features, such as Age, Body Mass Index, Lifestyle habits and many other factors. By using data from an established monocentric medical registry in Zürich, Switzerland, that has been in place for over 17 years, we are aiming to develop a supervised machine learning model using logistic regression. The goal is to have a model with an accuracy of at least 80% for the validation dataset. Strategies to deal with over and underfitting, missing data, as well as skewed variables are presented and discussed.

Table of Contents

1 Project Objectives

In orthopedic surgery and most likely in other domains of surgery, the surgical result is not only dependent on the surgeon's skill, but also on many factors beyond the surgeon's control. But how is surgical outcome measured? There are several metrics to gauge surgical success, such as objective measures like function and strength, or subjective measures for example level of pain, ability to do activities of daily living or patient satisfaction. The latter is arguably the most important one.

The goal of this project is to create a tool or an application for surgeons to use in their daily practice, whenever they have a patient with an indication for shoulder arthroplasty. The tool will predict patient satisfaction two years after shoulder arthroplasty based on a collection of preoperative factors, such as demographics, lifestyle habits, medical history and other attributes related to the surgery. By having a tool like that at hand, surgeons can manage the patients' expectations of the procedure or even encourage them to proceed, depending on the prediction. Maybe some patients are better off not having the surgery at that moment. By identifying the main contributors to patient satisfaction, targeted improvements in these areas can be achieved, ultimately leading to increased patient satisfaction and overall quality of life. The reason patient satisfaction two years postop is so important is, because at that time, the patients are usually very near to or at the final outcome of their surgery and are thus able to give an unbiased rating of their procedure and the outcome of which.

2 Methods

I will most likely use a local installation of python (Anaconda) on a private machine with the following libraries:

Pandas (pandas development team 2020)

Numpy (C. R. Harris et al. 2020)

Matplotlib.pyplot (Hunter 2007)

Seaborn (Waskom 2021)

Scikit-learn (Pedregosa et al. 2011)

I will be using a supervised machine learning model because I already have a dataset with a known outcome (patient satisfaction), and I am trying to predict unseen data. I am planning to begin with a linear regression model to predict patient satisfaction. First, I'll spit the data into training and validation datasets with an 80:20 ratio. Then, I'll train the model using the training dataset. After training, the trained model is fed the validation dataset. If the performance of the model with the validation dataset is sufficient, I have reached my goal. Else, I might try to use a different, better suited model. Or maybe I will need to adjust some parameters, depending on the performance of my model. I however need to keep in mind, that doing too much tuning of the can lead to overfitting, where the model is fitted too closely to the training data, such that potential randomness/noise in the training data is viewed as real patterns in the data. This leads to the model performing poorly on unseen data. It is not expected that the validation dataset performs as good or better than the training dataset. However, if the difference in performance between training and validation data is too big, this could be an indicator of overfitting.

3 Data

For this project, data from Schulthess Clinics shoulder arthroplasty registry (Marzel et al. 2020) will be used. The registry has been established in 2006 and short of 4000 shoulder arthroplasties have been documented since. There are about 220 shoulder arthroplasties being implanted per year at Schulthess Clinic. There have been unfruitful attempts to include patients from other clinics in the registry in the past, but to this day, the registry has remained monocentric.

For most patients, the data collection process is as follows: A few days prior to each shoulder arthroplasty procedure, the surgeon measures range of motion, strength and performs several clinical tests related to the patients' shoulder function. Additionally, preoperative imaging is evaluated and documented using several clinical scores. Before surgery, every patient is instructed to fill out a questionnaire to obtain subjective measures of pain and function. After surgery, the surgeon documents exactly what was done during the procedure such as the surgical approach, what kind of implants were used, additional treatments, adverse events etc. Postoperatively, the same parameters (except for surgical details) are documented at 6 months, 2 years, 5 years, 10 years, and 15 years. In the case of an Adverse Event, that is also documented, for example the severity of the adverse event, why it happened, if the issue was related to the procedure, what measures were applied to resolve the issue and what the outcome of these measures was. At each postoperative follow up, the patients are the same objective and subjective measures as before the surgery. See Figure 1 for a graphical overview of the patient journey related to their shoulder arthroplasty. Additionally, postoperatively the patients are asked how satisfied they are with their treatment. The question is phrased as *«Wurden ihre Erwartungen an die Operation erfüllt?».* The answer is a numeric rating scale from zero to ten, with zero being *«not at all»* and ten being *«completely»* (Figure 2).

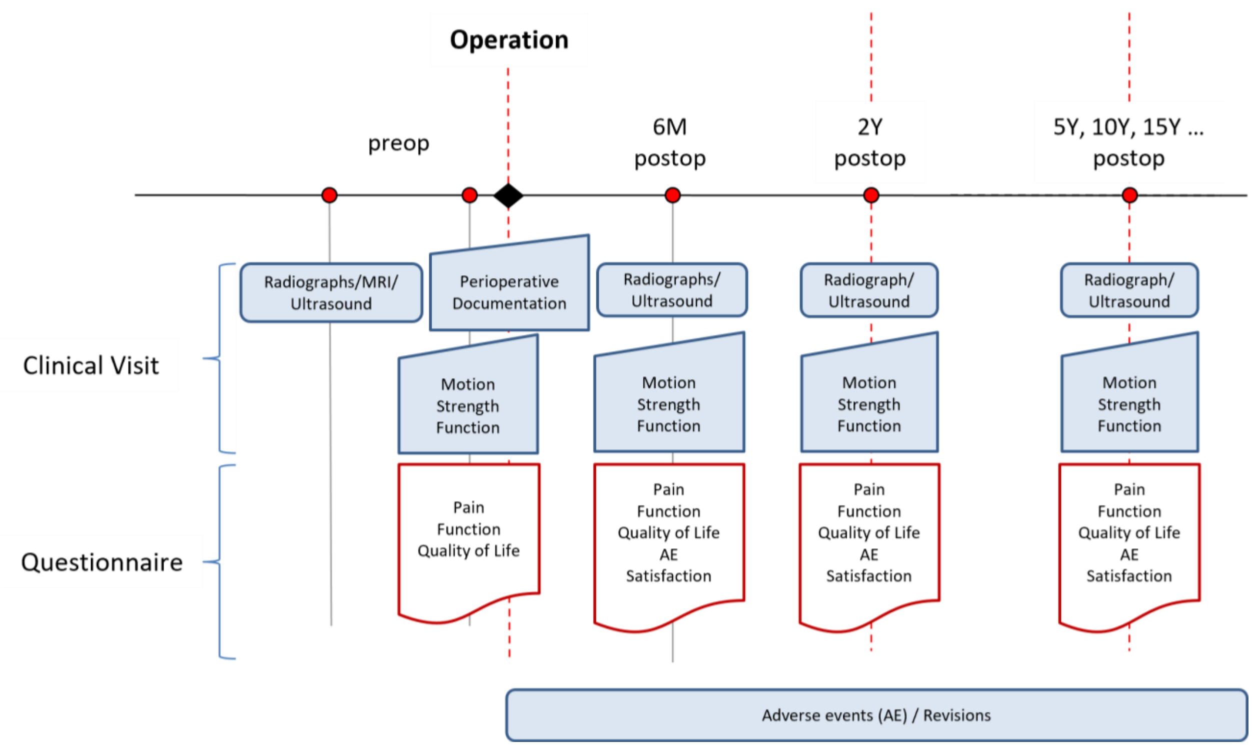


Figure 1: Timeline of Data Collection

Figure 2: Extract of the patient questionnaire with the outcome question

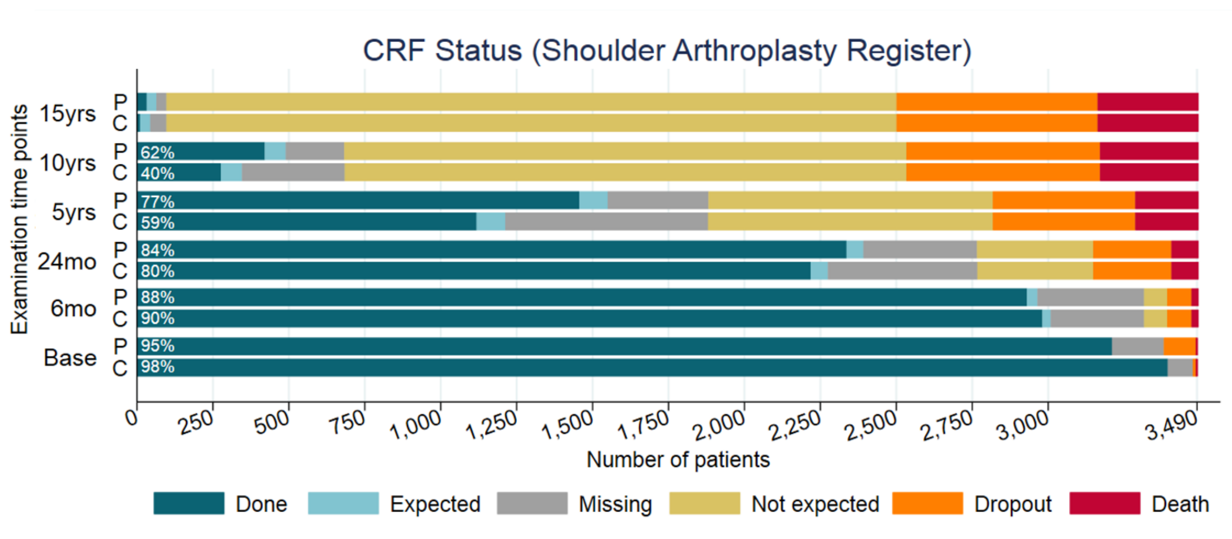


Figure 3: Case Report Form Status (C: Clinical Examination, P: Patient Questionnaire) per timepoint

What is important to note here is, that not the patient is followed, but the implant. Let's say a patient must undergo revision surgery. As soon as anything is changed with the implant (e.g. change of humeral component), we open a new record in the database, mark the old case as a dropout and only follow up the new case. The old and the new case are matched using the Case ID in our clinic information system. This way the whole "journey" of each patient is available. In Figure 3 the case report form (CRF) status by timepoint is depicted. The CRF status can take values from *Done* to *Death (Table 1).* The follow up rate shown on the left side of each bar in Figure 3 is calculated as follows*:*

$$\frac{nDone}{nExpected + nMissing}$$

CRF status	Description
Done	The patient questionnaire has been filled in or the clinical examination took place
Expected	The patient is **in the time range** for clinical examination/questionnaire
Missing	The patient is **beyond the time range**, but there is no data (various reasons)
Not expected	The patient is **not in the time range yet** for clinical examination/questionnaire
Dropout	The patient either doesn't want to participate anymore or has received a revision surgery
Death	The patient died

Table 1: A description of each CRF Status

Naturally, as time after surgery passes, the follow up rate declines. This can be explained in part by the fact that many patients are already in their 70s/80s when they receive shoulder arthroplasty (see Figure 4). Many patients also suffer from comorbidities, thus living in a nursing home or even death is not that unlikely. Also, as more time passes after surgery, revision surgery is getting more likely as the artificial implants don't last for eternity. This may lead to an eventual revision surgery, which results in a lower follow up rate, as only the new, revised implant will be followed up.

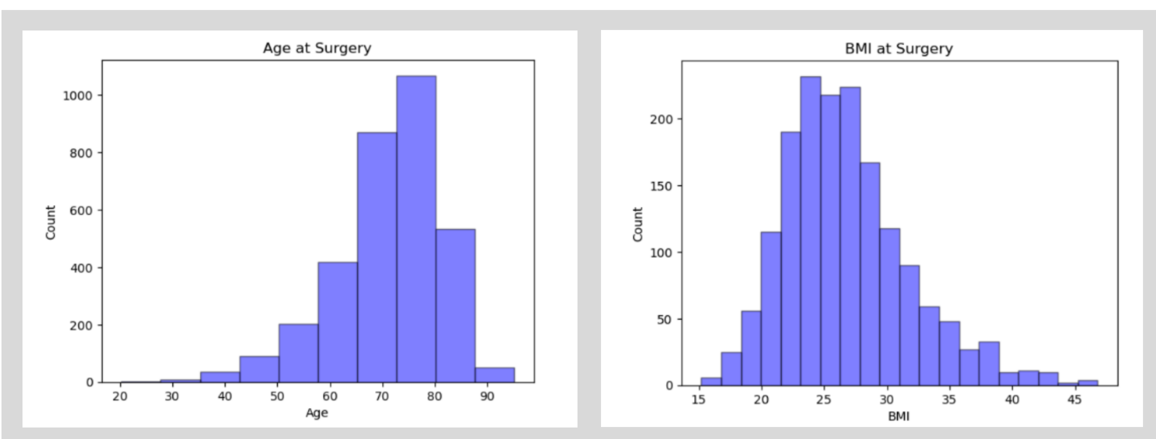


Figure 4: Histogram of Age/BMI at Surgery

All the collected patient data will be stored and secured using a state of the art electronic data capture system (EDC), namely REDCap (**R**esearch **E**lectronic **D**ata **Cap**ture) (P. A. Harris et al. 2019). It is an EDC created by Vanderbilt University, with useful features such as an audit trail, which allows for retrospective analysis of any changes in the data (who changed what and when). There is also a need for every user to have their own user account, which is secured by two-factor authentication when accessing REDCap from outside the clinics network. In Figure 5, an extract of the dataset that will be used is presented.

	gender	anshside	language	asa	sgadmsntyp	sginsurance	andiagn_new	anvoropyn	sgaplbrand	dominantside	ansmoking	analcohol	Age	bmi
1073	0.0	1.0	1.0	2.0	1.0	1.0	7.0	0.0	7.0	1.0	0.0	0.0	53.8	35.9
1093	0.0	1.0	1.0	3.0	1.0	1.0	1.0	0.0	2.0	1.0	0.0	1.0	70.0	40.5
1098	0.0	2.0	1.0	3.0	1.0	1.0	1.0	0.0	2.0	3.0	0.0	1.0	75.7	30.8
1099	1.0	2.0	1.0	3.0	1.0	1.0	7.0	1.0	7.0	1.0	0.0	1.0	62.6	39.4
1102	0.0	1.0	3.0	3.0	1.0	1.0	1.0	0.0	6.0	1.0	0.0	0.0	78.8	42.4

Figure 5: Extract of the dataset

4 Data Quality

A big challenge for the dataset at hand is missing data. First, the question is, are the values missing at random or is there a pattern in the missing values, which could lead to a biased model. I will try to fill in the missing values with the correct imputation method for this case. This could mean imputing the median or mean of the non-missing values.

Another challenge for this dataset is the outcome variable, which is heavily skewed to the right side (see Figure 6). Most patients are very satisfied with their treatment, and very few patients are highly dissatisfied with their treatment. What are strategies to deal with this heavily skewed outcome variable? The Oversampling Technique (Viloria, Lezama, and Mercado-Caruzo 2020) could be used to artificially increase the number of dissatisfied patients. Another option would be to combine classes at the cost of having bigger increments between groups but gaining more samples per class.

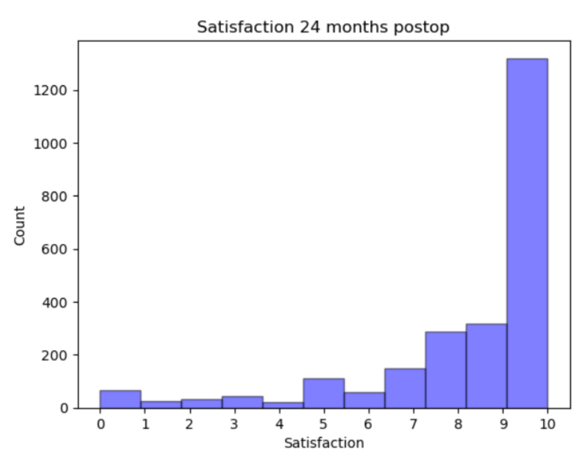
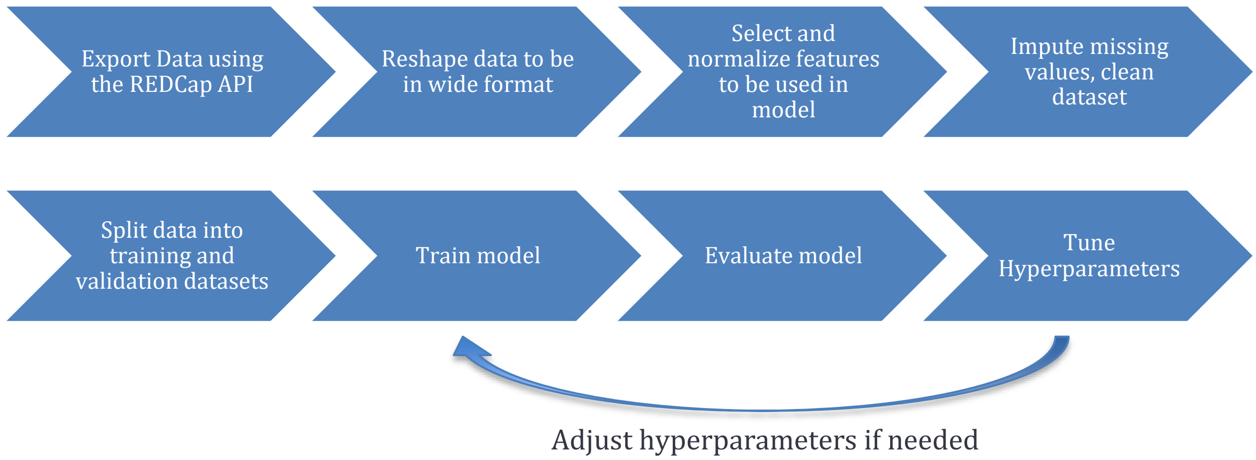


Figure 6: Histogram of the Patient Satisfaction 24 months postop

5 Data Flow



The first step is to export the data from REDCap using the REDCap API, which makes it very easy to access the latest version of the data. Then I need to reshape the data from long to wide, because initially, each timepoint is stored in a separate row. Because I need preoperative and postoperative data in the same row, a reshape is necessary. After reshaping, the relevant features must be selected. I chose features, that I thought would have the biggest impact on patient satisfaction. I could have performed a logistic regression first, and only include the features, which turned out statistically significant. After the feature selection, the features have to be normalized, to ensure the variability of a certain parameter is not based on the scale which. The next step is to clean the dataset, impute or omit missing values and deal with outliers. When it comes to the actual machine learning part, the first step is to split the data into training and validation datasets to be able to train the model with the training data and then evaluate the performance of the model with previously unseen data, i.e., the validation dataset. If the performance of the model is sufficient (80% accuracy) the job is done. If the performance is not good enough, some tuning of hyperparameters must be done and a retraining and subsequent evaluation of the model as well. The last two steps must be repeated as often as needed until the model reaches the desired performance.

6 Data Model

When it comes to the data model at the conceptual level, the goal is to develop an objective tool to predict patient satisfaction after shoulder arthroplasty based on certain preoperative features.

At the logical level, probably a logistic/linear multiclass regression model will be used to predict patient satisfaction with the preoperative features listed below.

Features:

- Age at surgery
- Sex
- Diagnosis
- ASA classification
- BMI
- Dominant side y/n
- Smoking y/n
- Alcohol consumption y/n
- Accident-related condition y/n
- Chief surgeon y/n
- Revision surgery y/n
- Reverse arthroplasty y/n
- Previous surgeries on the shoulder y/n
- Private insurance y/n

On the physical level, no additional infrastructure other than a conventional computer is needed as the model shouldn't require huge computational power. If it turns out the model is a bit more computationally heavy, switching from CPU to GPU could improve performance. If all fails, as a last resort, either buying a more powerful computer or renting a cloud-based GPU could help.

7 Risks

There is a risk that there are too many missing values not missing at random and I'm not able to address them in an appropriate manner, leading to bias. There is a chance of overfitting the data to the patient population at Schulthess Clinic, which means that the model may not necessarily be applicable to patients outside of Schulthess Clinic. To prevent this, I have to try to leave out those features, that don't contribute much to the variation in the data, but even then, there might be only poor generalization of the model.

8 Conclusions

A tool that can predict postoperative patient satisfaction can be a very powerful asset for every surgeon in the decision-making process of whether to perform surgery. I think, the requirements to build such a tool by applying machine learning are there, at least from a data point of view. If it really is possible to implement such a model and the performance of which remain unclear, however.

Acknowledgements

I'd like to thank all the people at Schulthess clinic that contributed to this dataset over the last 17 years, without their help this project wouldn't be possible. These people include all the members of the shoulder research team, the shoulder surgeons, and their secretaries.

Statement

„Ich erkläre hiermit, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen benutzt habe. Alle Stellen, die wörtlich oder sinngemäss aus Quellen entnommen wurden, habe ich als solche gekennzeichnet. Mir ist bekannt, dass andernfalls die Arbeit als nicht erfüllt bewertet wird und dass die Universitätsleitung bzw. der Senat zum Entzug des aufgrund dieser Arbeit verliehenen Abschlusses bzw. Titels berechtigt ist. Für die Zwecke der Begutachtung und der Überprüfung der Einhaltung der Selbstständigkeitserklärung bzw. der Reglemente betreffend Plagiate erteile ich der Universität Bern das Recht, die dazu erforderlichen Personendaten zu bearbeiten und Nutzungshandlungen vorzunehmen, insbesondere die schriftliche Arbeit zu vervielfältigen und dauerhaft in einer Datenbank zu speichern sowie diese zur Überprüfung von Arbeiten Dritter zu verwenden oder hierzu zur Verfügung zu stellen.“

Date: Signature(s):

References and Bibliography

Harris, Charles R, K Jarrod Millman, Stéfan J van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, et al. 2020. "Array Programming with {NumPy}." *Nature* 585 (7825): 357–62. https://doi.org/10.1038/s41586-020-2649-2.

Harris, Paul A., Robert Taylor, Brenda L. Minor, Veida Elliott, Michelle Fernandez, Lindsay O'Neal, Laura McLeod, et al. 2019. "The REDCap Consortium: Building an International Community of Software Platform Partners." *Journal of Biomedical Informatics*. https://doi.org/10.1016/j.jbi.2019.103208.

Hunter, J D. 2007. "Matplotlib: A 2D Graphics Environment." *Computing in Science & Engineering* 9 (3): 90–95. https://doi.org/10.1109/MCSE.2007.55.

Marzel, Alex, Hans Kaspar Schwyzer, Christoph Kolling, Fabrizio Moro, Matthias Flury, Michael C. Glanzmann, Christian Jung, et al. 2020. "The Schulthess Local Shoulder Arthroplasty Registry (SAR): Cohort Profile." *BMJ Open* 10 (11). https://doi.org/10.1136/bmjopen-2020-040591.

pandas development team, The. 2020. "Pandas-Dev/Pandas: Pandas." Zenodo. https://doi.org/10.5281/zenodo.3509134.

Pedregosa, F, G Varoquaux, A Gramfort, V Michel, B Thirion, O Grisel, M Blondel, et al. 2011. "Scikit-Learn: Machine Learning in {P}ython." *Journal of Machine Learning Research* 12: 2825–30.

Viloria, Amelec, Omar Bonerge Pineda Lezama, and Nohora Mercado-Caruzo. 2020. "Unbalanced Data Processing Using Oversampling: Machine Learning." In *Procedia Computer Science*. https://doi.org/10.1016/j.procs.2020.07.018.

Waskom, Michael L. 2021. "Seaborn: Statistical Data Visualization." *Journal of Open Source Software* 6 (60): 3021. https://doi.org/10.21105/joss.03021.