MEDINFO 2021: One World, One Health – Global Partnership for Digital Innovation
P. Otero et al. (Eds.)

doi:10.3233/SHTI220186

Using Machine Learning on Imbalanced Guideline Compliance Data to Optimize Multidisciplinary Tumour Board Decision Making for the Management of Breast Cancer Patients

My-Anh Le Thien[a], Akram Redjdal[a], Jacques Bouaud[b,a], Brigitte Séroussi[a,c,d]

[a] *Sorbonne Université, Université Sorbonne Paris Nord, Inserm, UMR S_1142, LIMICS, Paris, France*
[b] *AP-HP, DRCI, Paris, France*
[c] *AP-HP, Hôpital Tenon, Paris, France*
[d] *APREC, Paris, France*

Abstract

Complex breast cancer cases that need further multidisciplinary tumor board (MTB) discussions should have priority in the organization of MTBs. In order to optimize MTB workflow, we attempted to predict complex cases defined as non-compliant cases despite the use of the decision support system OncoDoc, through the implementation of machine learning procedures and algorithms (Decision Trees, Random Forests, and XGBoost). F1-score after cross-validation, sampling implementation, with or without feature selection, did not exceed 40%.

Keywords:

Supervised Machine Learning, Breast Cancer, Clinical Decision support systems

Introduction

Multidisciplinary tumour boards (MTBs) are a staple for managing cancer patient treatment. Patient-specific treatment plans as recommended by clinical practice guidelines (CPGs) have improved patients' outcomes and clinicians are encouraged to follow them [1].

Oncodoc is a guideline-based decision support system developed to provide patient-specific recommendations and promote CPG implementation for breast cancer management. When OncoDoc was routinely used in MTBs at the Tenon hospital (Paris, France), CPG compliance reached 91.7% for invasive breast cancer [2].

Thus, despite the use of Oncodoc, non-compliant cases were observed. We consider those cases as "complex cases", since they are not standard enough to be covered by CPGs and consequently should have priority in MTB discussions.

Our objective is to implement machine learning methods to identify patients not handled by CPGs and optimize patient triage ahead of MTBs. This way, non-complex cases might be treated faster with the support of a CDSS, and more time could be allocated to complex cases. We tested a wide selection of models and sampling techniques and compared their performance with recall and precision to select the most efficient combination.

Methods

Data was collected from the existing OncoDoc database, and included MTB decisions for adult women treated for breast cancer from February 2007 to September 2009. Data consisted of 1,887 MTB decision instances (1,054 patients) with 127 collected variables. A sizable amount of variables was incomplete due to Oncodoc's architecture as a decision tree: non-relevant data to the case is not asked, and therefore not entered.

We applied supervised machine learning, with labelled training datasets and all values predicted from the test datasets verified against the actual class. Cases where clinicians did not comply with OncoDoc recommendations were labelled as "*complex*" and cases where OncoDoc recommendations were followed by MTB clinicians as "*non-complex*".

Whenever possible, missing values were assigned so as to remain logically sound, e.g., when a tumour was non-invasive, all tumour-invasive-related variables were filled as "not applicable". All missing or "not applicable" values were also considered as integers: excision margins outlying invasive tumour were originally coded as invaded (1) or not (0) and were eventually encoded as *Missing* (0), *Not applicable for non-invasive tumour* (1), *Non-invaded* (2) and *Invaded* (3). Additional variables were built to reflect factors of clinical complexity known from the literature, e.g., triple negative breast cancer patients (hormonal receptors = negative AND Her2 = negative).

The final processed dataset is comprised of 1,887 instances and 70 variables.

The following standard procedures were applied:

- Stratification on complexity to keep the original dataset class ratio in the train-test datasets.
- With k-fold cross-validation: the training set is split into k sets (here, k=5). Each of the k folds is used as test set against the rest, and model performance is given by the averaged scores.

Since data was severely imbalanced with few complex cases (7.5%), we tested the following samplers on each trainset generated for cross-validation, in order to offset data imbalance: random under sampling (RUS), random over sampling (ROS), SMOTE, ADASYN, and SMOTE and Edited Nearest Neighbours (SMOTEEN) [3].

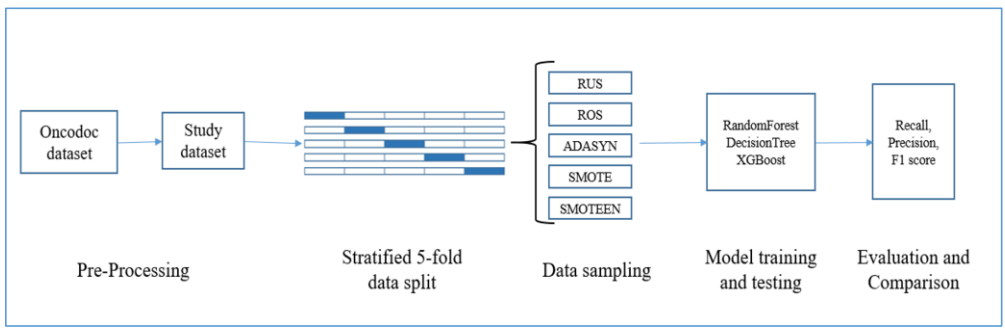

Figure 1– Analytic plan

The following classification models were trained and tested: Decision Trees (DT), Random Forests (RF) and XGBoost (XGB) [4, 5]. We narrowed down the best value for each model's hyperparameters using Random Search and Grid Search. Random Search finds a broad range of optimized parameters after testing random combinations of parameters and Grid Search then narrows down the hyperparameters through systematic testing. We then selected the most useful variables for each model using feature selection.

We used precision, recall, and F1-score for model evaluation. Accuracy was considered to be unreliable as the test set was unbalanced (if 90% of data belong to class A, a model might simply choose to systematically class data as class A to obtain a 90% accuracy). We compared the mean cross-validation scores for each sampler/model combination. (cf. Figure 1).

Results

Decision Trees and Random Forests presented best F1 score with the ROS sampler (DT: 37.4%, RF: 33.5%), and XGBoost performed best with the RUS sampler (38.1%). XGB presented overall better results than DT and RF but was still insufficient for an exploitable model. (cf. Table 1)

Table 1 – F1-score by model and sampling technique

Model	DT	RF	XGB
No sampling	28.0%	31.5%	36.9%
RUS	24.3%	29.0%	38.1%
ROS	37.4%	33.5%	37.6%
ADA	34.4%	29.9%	37.6%
SMOTE	32.0%	28.5%	37.1%
SMOTEEN	28.3%	29.5%	37.1%

DT: DecisionTrees, RF: RandomForests, XGB: XGBoost

Discussion and conclusion

Machine learning approaches did not yield a model efficient enough for classifying complex cases as F1-scores were unsatisfactory. We tested multiple samplers to correct data imbalance with varying results, as one sampler might improve one model's performance and worsen another. Hyperparameters tuning and feature selection improved all models but remains insufficient.

We defined complexity as non-conformity with OncoDoc and acknowledge the possible limits associated: complex cases which eventually were compliant with OncoDoc after a lengthy discussion are not identified in our study. Likewise, some treatments might be dismissed without posing difficulty to MTB (patient's preference). Further analyses and reviews of MTB decisions might give us a more accurate definition of complexity for further studies.

We tested several algorithms, used samplers for data imbalance and optimized hyperparameters and feature selection but models were not efficient enough to classify complex patients in the Oncodoc dataset. Different models, additional data and/or data structuring might improve results.

References

[1] Kreienberg R, Wöckel A, Wischnewsky M (2018) Highly significant improvement in guideline adherence, relapse-free and overall survival in breast cancer patients when treated at certified breast cancer centres: An evaluation of 8323 patients Breast. 40:54–59.

[2] B. Séroussi, C. Laouénan, J. Gligorov, S. Uzan, F. Mentré, and J. Bouaud, Which breast cancer decisions remain non-compliant with guidelines despite the use of computerised decision support?, *Br J Cancer*. **109** (2013) 1147–1156. doi:10.1038/bjc.2013.453.

[3] Sun Y, Wong AKC, Kamel MS (2009) Classification of imbalanced data: a review Int J Patt Recogn Artif Intell. 23:687–719.

[4] C. Kingsford, and S.L. Salzberg, What are decision trees?, Nat Biotechnol. 26 (2008) 1011–1013. doi:10.1038/nbt0908-1011.

[5] R. Santhanam, N. Uzir, S. Raman, and S. Banerjee, Experimenting XGBoost Algorithm for Prediction and Classification of Different Datasets, in: 2017.

Address for correspondence

My-Anh LE THIEN
LIMICS U1142,
15, rue de l'école de médecine,
75006 Paris, France
myanh.lethien@gmail.com