

## Supervised Learning Methods to Predict Outcomes in Orthopaedic Research

Federico Solla<sup>1\*</sup>, Domenico Bertoncelli<sup>2</sup>, Virginie Rampal<sup>1</sup>, Audrey Voury-Pons<sup>1</sup> and Carlo M Bertoncelli<sup>1,3</sup>

<sup>1</sup>*Pediatric Orthopaedic Surgery, Lenval University Children Hospital, Nice, France*

<sup>2</sup>*Department of Information Engineering, Computer Science and Mathematics, University of LAquila, Italy*

<sup>3</sup>*Institut Henri Germain, Lenval Foundation, Nice, France*

\***Corresponding Author:** Federico Solla, Pediatric Orthopaedic Surgery, Lenval University Children Hospital, Nice, France.

**Received:** December 18, 2017; **Published:** January 06, 2018

### Abstract

**Aim:** To propose supervised learning methods named “PredictMed” based on the statistical learning theory in clinical research.

**Methods:** An algorithm is built to predict a clinical outcome, e.g. postoperative complications yes/no. After the choice of independent variables, a bespoke multiple regression model is built to predict the probability for a patient to reach the outcome. The originality of PredictMed is that the regression model is first applied on a “training set” of patients (80% of the sample) and then on a “test set” (the remaining 20%) in order to assess the effectiveness in predicting the outcome for a new patient. The performance of the bespoke model is thus evaluated by measuring and averaging its accuracy, specificity and sensitivity on several different test sets of patients.

**Conclusions:** PredictMed can predict outcomes in large series of patients ( $n \geq 70$ ). The evaluation of performance is more intuitive than for “usual” multiple regressions. Although its application requires advanced statistical competence, this method is easily intelligible in articles since it provides understandable values, e.g. accuracy, specificity and sensitivity, allowing clear clinical applications. sets of patients.

**Keywords:** *Supervised Learning Methods; PredictMed; Test Set*

### Background

Machine learning is a contemporary branch of Statistics and Artificial Intelligence particularly adapted for analysis of complex data [1]. It combines many parameters and constructs an optimized mathematical equation to predict an endpoint with the least error [2].

Supervised models based on the statistical learning theory by Vapnik and Chervonenkis [3,4] are better suited to predict to which class (or group) a new patient belongs if the class of each patient is known a priori. This methodology has been recently used for cancer classification, novel drug discovery, genomics, risk scoring, patients’ stratification, diagnostic tools [5-11], surgery [12-15], and disability [16,17].

To our knowledge, few articles report on this method in surgical [12,13] or orthopaedic researches [14]. Sing., et al. used machine learning to classify recent literature on spine [15].

Our aim is to propose a supervised learning method, named “PredictMed”, to analyze large series of patients’ data with multiple independent parameters in clinical research. PredictMed has been developed [16] and validated [17] for predicting neuromuscular scoliosis in adolescents with cerebral palsy. This method can predict outcomes in prognostic, diagnostic, case-control, cohort, cross-sectional and prospective studies.

### Methods

An algorithm is built to predict the probability of a clinical outcome, e.g. postoperative complication. As examples, we will present logistic regressions with binary endpoints (yes/no). Moreover, this approach is possible for quantitative (e.g. Cobb angle) or polynomial qualitative outcomes (e.g. right/ left/ both), which require respectively linear regressions or polynomial logistic regressions [18-20].

The first steps are similar to “usual” multivariate analysis and will not be detailed. Independent variables are chosen from literature review and personal opinion. They are first analyzed with univariate tests (e.g. T-tests, Fisher’s exact test), and then used in a bespoke multiple regressions to evaluate the probability for a patient to reach the binary endpoint. This is feasible with open source software ‘R’, using the General Linear Model ‘glm()’ function [21,22]. The originality of “PredictMed” is that the multi-regression is first applied on a “training set” of patients (80% of the sample) and then on a “test set” (the remaining 20%) in order to assess the effectiveness in predicting the endpoint for a new patient. Ideally, each value should be present in each test set; for this, the sample should be large enough and/or the values of the outcome (yes/no) should be almost equally distributed (empirically, from 1:1 up to 3:1).

Thus, the model predicts for each subject the probability (having by definition  $0 < \text{Probability} < 1$ ) for him/her to reach the outcome, given the values of the independent variables.

In order to minimize the influence of Training and Test Sets (TTS) composition, we can use a cross validation, by randomly generating multiple different couples of TTS and calculating specificity, sensitivity and accuracy (SSA) in prediction for each couple of TTS. The performance of the model is then evaluated by averaging its SSA on several different TTS.

Sensitivity is defined as the proportion of true positives (TP). Specificity is defined as the proportion of true negatives (TN). Accuracy is the precision of the measurement system, related to reproducibility and repeatability. It is calculated as follows:  $(TP+TN) / \text{the whole sample}$ .

We usually implement it by using the predict glm() R-function (a). We calculate a probability in the form of  $P[\text{outcome} = \text{yes} \mid \text{glm}(\text{independent variables})]$  for each patient of each test set. If the probability is above a defined threshold, we can predict if each subject will reach the outcome “yes”. Each prediction is thus noted as true positive, true negative, false positive, false negative.

Therefore, the program provides multiple models with measured SSA for various combinations of independent variables and various thresholds of probability (Table 2). The threshold of probability can be chosen a priori, depending on the aim of the study. Otherwise, the scholar chooses a model (and fixes the corresponding probability threshold) among the others after the analysis (post hoc), depending on the calculated values of SSA to maximize sensitivity or specificity or to equilibrate their values, in accordance with the aim of the research (Table 2). Acceptable values of SSA in clinical research are  $\geq 0.70$  [23-25].

**Results: Presentation and Interpretation**

The number of subjects for TTS is enounced, as well as the number of couples of TTS. After the results of univariate tests, the predictive algorithm is presented as a multivariate analysis with a defined threshold for the probability and averaged SSA (Table 1).

Independent Variables	Logistic Regression for outcome Yes/No				
	Odds Estimate	Standard Error	Z Ratio	Prob (> z ) p Value	Value
Intercept	-13.1294	4.6271	-2.838	0.004547	**
A	0.8034	1.1997	0.670	0.503049	
B	-0.1718	0.8340	-0.206	0.836773	
C	-0.2692	0.5180	-0.520	0.603232	
D	-1.6755	1.2376	-1.354	0.175786	
E	2.2939	1.0756	2.133	0.032954	*
F	-3.3473	1.6499	-2.029	0.042484	*
G	6.0862	1.6487	3.692	0.000223	***

**Table 1:** Example of a multiple logistic regression.

\* < 0.05 \*\* < 0.01 \*\*\* < 0.001

Significant factors are E, F and G ( $p < 0.05$ ), adjusted on cofactors A, B, C and D ( $p > 0.05$ ). For every unit increase in E, the log odds =  $\ln(p/1-p)$  increases 2.2939 times (where  $p = \text{probability to have outcome YES}$ ). A similar interpretation is done for F and G. On the opposite, the influence of A, B, C and D is not significant.

The table 2 presents five models with different thresholds of probability and different independent variables, as selected by the program. In analyzing data, the first model should be chosen for higher sensibility (82% in predicting the outcome YES), whereas the second provides a better specificity. Nevertheless, the fifth one is the worst for all values. The definitive choice of one model should be discussed and motivated.

Variables	Probability Threshold	Accuracy*	Sensitivity*	Specificity*
A B C D E F G	0.4	<b>0.72</b>	<b>0.82</b>	0.63
A B C D E F G	<b>0.6</b>	0.71	0.75	<b>0.79</b>
A C D F G	0.5	<b>0.72</b>	0.72	0.64
A C D F G	0.6	0.70	0.67	0.65
A D G	0.5	0.68	0.64	0.62

**Table 2:** Example of various models as proposed by the program “R”.

Call: glm(formula = model.formula, family = binomial(logit), data = input\_file)

In bold: the highest values of each column

\*mean values

## Conclusions

Supervised learning statistics can predict outcomes in large series of patients. The performance is better evaluated for this kind of analysis than for “usual” multiple regressions. The application of these algorithms requires advanced competence in statistics and informatics. However, the results of PredictMed are easily intelligible once they are written in articles as they provide understandable values, i.e. accuracy, specificity and sensitivity, allowing intuitive clinical applications. In clinical research, the main focus of machine learning is not to replace the human decision-making but to provide help in complex situations and improve the quality of predictions [4].

## Take Home Points:

- Supervised learning statistics “PredictMed” can predict outcomes in large series of patients with a clear evaluation of the performance.
- The application of “PredictMed” is possible with “R” program and requires advanced competence in statistics and informatics.
- The results of “PredictMed” are easily intelligible because they provide understandable values, i.e. accuracy, specificity and sensitivity, allowing clear clinical applications.

## Disclosures

- F.S. received financial support for attending symposia and congresses from Medicea Int. unrelated to this work.
- VR, DB, CMB, AVP: no conflicts.
- This research did not receive any specific grant or funding.

## Bibliography

1. Hastie T., *et al.* “The Elements of Statistical Learning”. *Elements* 1 (2009): 337-387.
2. Cherkassy V., *et al.* “Model complexity control for regression using VC generalization bounds”. *IEEE Transactions on Neural Networks* 10.5 (1999): 1075-1089.
3. Vapnik V. “The nature of statistical learning theory”. Springer Science and Business Media (2013).
4. Mossotto E., *et al.* “Classification of Paediatric Inflammatory Bowel Disease using Machine Learning”. *Scientific Reports* 7 (2017): 2427.
5. Upstill-Goddard R., *et al.* “Support Vector Machine classifier for estrogen receptor positive and negative early-onset breast cancer”. *PLoS One* 8.7 (2013): e68606.
6. Mathé C., *et al.* “Current methods of gene prediction, their strengths and weaknesses”. *Nucleic Acids Research* 30.19 (2002): 4103-4117.
7. Kannel WB., *et al.* “Precursors of sudden coronary death. Factors related to the incidence of sudden death”. *Circulation* 51.4 (1975): 606-613.
8. Criminisi A. “Machine learning for medical images analysis”. *Medical Image Analysis* 33 (2016): 91-93.
9. Woodruff PG., *et al.* “T-helper type 2-driven inflammation defines major subphenotypes of asthma”. *American Journal of Respiratory and Critical Care Medicine* 180.5 (2009): 388-395.
10. Hu X., *et al.* “Review Artificial neural networks and prostate cancer--tools for diagnosis and management”. *Nature Reviews Urology* 10.3 (2013): 174-182.
11. Shouval R., *et al.* “Machine learning for prediction of 30-day mortality after ST elevation myocardial infarction: An Acute Coronary Syndrome Israeli Survey data mining study”. *International Journal of Cardiology* 246 (2017): 7-13.

12. Fard MJ, *et al.* "Automated robot-assisted surgical skill evaluation: Predictive analytics approach". *International Journal of Medical Robotics* (2017).
13. Donal E., *et al.* "Machine learning to support decision-making for cardiac surgery during the acute phase of infective endocarditis". *Heart* 103.18 (2017): 1396-1397.
14. Oh E., *et al.* "Prediction of pathologic femoral fractures in patients with lung cancer using machine learning algorithms: Comparison of computed tomography-based radiological features with clinical features versus without clinical features". *Journal of Orthopaedic Surgery* 25.2 (2017): 2309499017716243.
15. Sing DC., *et al.* "Machine Learning-Based Classification of 38 Years of Spine-Related Literature Into 100 Research Topics". *Spine (Phila Pa 1976)* 42.11 (2017): 863-870.
16. Bertonecelli CM., *et al.* "Risk factors for developing scoliosis in cerebral palsy: A cross sectional descriptive study". *Journal of Child Neurology* 32.7 (2017): 657-662.
17. Bertonecelli CM., *et al.* "Validation of a Clinical Prediction Model for the Development of neuromuscular Scoliosis. a Multinational Study". *Pediatric Neurology* (2017).
18. Kruppa J., *et al.* "Probability estimation with machine learning methods for dichotomous and multicategory outcome: theory". *Biometrical Journal* 56.4 (2014): 534-563.
19. Kruppa J., *et al.* "Risk estimation and risk prediction using machine-learning methods". *Human Genetics* 131.10 (2012): 1639-1654.
20. Baştanlar Y and Ozuysal M. "Introduction to machine learning". *Methods in Molecular Biology* 1107 (2014): 105-128.
21. The R Core Team. A language and environment for statistical computing. R Vienna, Austria: R Foundation for Statistical Computing (2012).
22. Hilbe J. "Logistic regression using R". Book Medical Publishers Chapman and Hall Book. Boca Raton FL USA (2009).
23. Silva WP, *et al.* "Accuracy, Sensitivity and Specificity of Fine Needle Aspiration Biopsy for Salivary Gland Tumors: A Retrospective Study from 2006 to 2011". *Asian Pacific Journal of Cancer Prevention* 17.11 (2016): 4973-4976.
24. Thombs BD and Rice DB. "Sample sizes and precision of estimates of sensitivity and specificity from primary studies on the diagnostic accuracy of depression screening tools: a survey of recently published studies". *International Journal of Methods in Psychiatric Research* 25.2 (2016): 145-152.
25. Wen Z., *et al.* "Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS® implementations". NESUG Proceedings: Health Care and Life Sciences, Baltimore, Maryland (2010): 1-9.

**Volume 9 Issue 1 January 2018**

**© All rights reserved by Teresa Bas., *et al.***