A decision-making tool to fine-tune abnormal levels in the complete blood count tests

Marta Avalos, Hélène Touchais, Marcela Henríquez-Henríquez

A decision-making tool to fine-tune abnormal levels in the complete blood count tests

Marta Avalos-Fernandez MARTA.AVALOS-FERNANDEZ@U-BORDEAUX.FR
SISTM team, INRIA BSO, F-33405, Talence, France
Univ. Bordeaux, INSERM, BPH U1219, F-33000 Bordeaux, France

Hélène Touchais HELENE.TOUCHAIS@U-BORDEAUX.FR
SISTM team, INRIA BSO, F-33405, Talence, France
INP-ENSEEIHT, F-31000, Toulouse, France

Marcela Henríquez-Henríquez MARCELA.HENRIQUEZH@INTEGRAMEDICA.CL
ELSA Clinical Laboratory, IntegraMedica, part of BUPA, Santiago, Chile
Cardio MR, Millennium Science Initiative Program, ANID, Santiago, Chile

Abstract

The complete blood count (CBC) performed by automated hematology analyzers is one of the most ordered laboratory tests. It is a first-line tool for assessing a patient's general health status, or diagnosing and monitoring disease progression. When the analysis does not fit an expected setting, technologists manually review a blood smear using a microscope. The International Consensus Group for Hematology Review published in 2005 a set of criteria for reviewing CBCs. Commonly, adjustments are locally needed to account for laboratory resources and populations characteristics. Our objective is to provide a decision support tool to identify which CBC variables are associated with higher risks of abnormal smear and at which cutoff values. We propose a cost-sensitive Lasso-penalized additive logistic regression combined with stability selection. Using simulated and real CBC data, we demonstrate that our tool correctly identify the true cutoff values, provided that there is enough available data in their neighbourhood.

Keywords: Interpretability, Lasso, GAM, Imbalance, Population Health.

1. Introduction

The complete blood count (CBC) with leukocyte differential count performed by automated hematology analyzers is one of the most ordered laboratory tests since used as first-line tool for assessing a patients general health status, diagnosing and monitoring disease progression and therapy. An automated hematology analyzer is a machine that rapidly counts and discriminates blood cells depending on its size, complexity or staining. When the analysis does not fit an expected setting, the machine triggers a warning flag. Then, technologists manually prepare and review a blood smear using a microscope. The manual technique is more efficient for the identification of cytological nuances, particularly in the presence of immature or abnormal cells. However, this is one of the most time-consuming hematology laboratory tests and requires a high degree of technical skill to minimize errors inherent in the subjectivity of the procedure. There is great interest in reducing the number of automated CBCs requiring manual blood smear reviews without sacrificing the quality of patient care (Clé, 2017).

The International Consensus Group for Hematology Review published in 2005 a set of rules as criteria for reviewing CBCs (Barnes et al., 2005). However, laboratories must adjust these general rules to their patients and machine characteristics. Commonly, adjustments are done accounting for the technical capabilities of hematology automatons, the population/sample distributions of the CBC data and permanently updated expert guidelines. The goal is to achieve the best trade-off between optimizing human and financial resources (i.e. minimizing the false positive rate) and optimizing the quality of patient care (i.e. minimizing the false negative rate, usually set in advance).

The manual smear is also considered as an internal quality control procedure for the evaluation of the automated hematology machine parameters. Regularly, all (flagged and unflagged) samples produced during a limited period of time are analyzed to determine the need for adjustments of the screening criteria for manual blood smear review. These large amounts of data regularly produced could be used to train machine learning algorithms that could be integrated to the quality control procedure as a decision-making tool to fine-tune abnormal levels in the CBC tests.

The objective of the present work was to provide a machine learning tool for decision support in fine-tuning abnormal levels in the CBC tests at laboratory-level. Explicitly, we aim to identify which CBC variables are associated with a higher risk of abnormal manual smear and at which cutoff values. This involves that the machine learning tool must provide interpretable results. However, that should surely not be an argument for sacrifying predictive performance.

Our proposal is based on cost-sensitive Lasso-penalized additive logistic regression. Additive functions are considered to belong to the space of piecewise constant functions. The natural sparsity encouraged by the Lasso penalty is combined with a stability selection procedure to enforce model stability (Meinshausen and Bühlmann, 2010). The rationale behind this choice is the following. Manual smear results are the binary outcome (abnormal/normal). CBC tests data are the quantitative predictors. The logistic regression model is a popular choice in the biomedical field. Categorisation of continuous variables is common in clinical studies since results are easier to interpret. Thus, we apply additive models fitted by piecewise constant polynomials. Once the piecewise constant basis functions have been set, we have gone through in a linear space setting.

Now, a key question is how to determine the optimal number and location of cutoff points. Standard variable selection procedures are applied when only one or a couple of quantitative predictors are present (Liquet and Riou, 2019; Barrio et al., 2013). Since the CBC data consist in some twenty biological measures, multiplied by the number of basis functions, the Lasso appears as a valuable alternative (Tibshirani, 1996). To encourage stable results we used Stability selection: only variables frequently selected by the Lasso analysis over data subsamples are retained (Meinshausen and Bühlmann, 2010).

Finally, normal manual smear blood tests are much more usual than abnormal tests (less than 10% in the present study). We are faced with an imbalanced learning problem. The two main families of methods in imbalanced binary classification are sampling and cost-sensitive methods (He and Garcia, 2009). A simple cost-sensitive strategy consists in weighting individuals' contributions to the likelihood function to account for the degree of disequilibrium or importance (Dmochowski et al., 2010).

2. Method

Let $\{(X_{i1}, \ldots, X_{ip}, Y_i)\}_{i=1}^n$ be an i.i.d. n-sample and $\{(x_{i1}, \ldots, x_{ip}, y_i)\}_{i=1}^n$ a realization of the sample. Y is binary (0 for normal, 1 for abnormal smears). Let us note $\mathbf{x}_j = (x_{1j}, \ldots, x_{nj})^\top$. Consider the additive logistic model: $\text{logit}(P(Y_i = 1|x_{i1}, \ldots, x_{ip})) = \beta_0 + \sum_{j=1}^p f_j(x_{ij})$, where $f_j : \mathbb{R} \to \mathbb{R}$ are unknown centred functions and β_0 the intercept. Let $\{\chi_j^k(\mathbf{x}_j)\}_{k=1}^{p_j}$ be a fixed basis of functions and denote $\boldsymbol{\chi}_j^k = \chi_j^k(\mathbf{x}_j)$. Any f_j can be expanded in terms of these basis and the unknown parameters $\{\beta_j^k\}_{k=1}^{p_j}$: $f_j(\mathbf{x}_j) = \sum_{k=1}^{p_j} \beta_j^k \boldsymbol{\chi}_j^k = \boldsymbol{\chi}_j \boldsymbol{\beta}_j$. Consider now the particular case of piecewise constants $\chi_{ij}^k = 1$ if $x_{ij} > q_j^k$ and 0 otherwise, with q_j^k the k-th value of a collection of K_j fixed values. Considering the negative log-likelihood loss function, we can write the weighted penalized optimization problem as:

$$\max_{\boldsymbol{\beta}} \left\{ \sum_{i=1}^n \omega_i \ln \frac{e^{y_i \sum_{j=1}^p \boldsymbol{\chi}_{ij} \boldsymbol{\beta}_j}}{1 + e^{\sum_{j=1}^p \boldsymbol{\chi}_{ij} \boldsymbol{\beta}_j}} - \lambda \|\boldsymbol{\beta}\|_1 \right\}, \tag{1}$$

where the weights $\omega_i = \omega > 1$ if $y_i = 1$ and 1 if $y_i = 0$ are used to account for the degree of imbalance in the minority class. $\lambda > 0$ controls regularization. The intercept is omitted. The selection probability is computed and used as a continuous measure of the stability associated to the Lasso estimates $\widehat{\beta_j^k}$.

Amato et al. (2016) conducted a comprehensive review of parsimonious additive models and reformulated the estimation problem in terms of group Lasso. Sparse group Lasso and overlapping group Lasso are alternative reformulations (Chouldechova and Hastie, 2015; Lou et al., 2016). Our proposal is close to these approaches. Nevertheless, by choosing piecewise constant functions, which conveniently categorises CBC data, interpretation, estimation, and computational issues are dramatically simplified. The fused Lasso has also been adapted to multiple change-point detection by piecewise constant functions (Petersen and Witten, 2019). However, this approach is not well suited for large data.

2.1. Evaluation

CBCs and blood smear reviews for 9 594 patients performed at the clinical laboratory of the Pontificia Universidad Catolica de Chile in 2016 were available. CBC data consisted in hemoglobin (Hg in g/dL), hematocrit (Ht in %), mean corpuscular hemoglobin concentration (MCHC in g/dL), mean corpuscular volume (MCV in fL), erythrocytes (Er in $10^6/\mu L$), platelets (P in $10^3/\mu L$), red blood cell distribution width (RDW-CV in %), leukocytes (Le in $10^3/\mu L$), immature granulocytes (IG in %) and the leukocyte differential count which includes neutrophils (N), basophils (B), eosinophils (Eo), monocytes (M) and lymphocytes (Ly) (in $10^3/\mu L$ and %). Alarms of suspected alterations of blood cells (binary), sex (binary) and age (in years) were also reported. The response, normal/abnormal smear, is imbalanced: only 7% of the smears were abnormal.

Real data served as the core of a simulation study to evaluate if our procedure was able to detect relevant predictors and relevant cut-off values. The features were defined by the original data. Y was generated assuming the logistic model. We considered 4 scenarios: A and B, with only 2 (low-correlated) relevant predictors and C and D with 5 relevant predictors which shared moderate to high correlation between them and low correlation with all the other variables. Only one β_j^k per variable was different to 0. In scenarios A and D, they were placed in the extreme percentiles, where observations are scarce. Inversely, in scenarios B and C, they were placed in frequently observed values. β_0 was calibrated to achieve 7% of events. Y was generated 100 times and selection probabilities computed.

3. Results

In all the 4 scenarios, irrelevant variables were infrequently selected (Figure 1). In scenarios A and D, in which the true features have to be learn from scarce examples, their variability is slightly higher and the target value for the true cut-off values are less often achieved. Indeed, in scenario A, the 2 true values are identified as non zero in 52% and 68% of the situations, respectively. In scenario D, the 5 true values are identified as non zero in few situations. The procedure doesn't present good identification performance when relevant predictors are correlated to other relevant predictors and the true cut-off values are placed in the extreme percentiles. Inversely, in scenarios B and C, the capability to detect the real cut-off values is excellent. Values in the neighbourhood of the real cut-off values are most frequently selected than other irrelevant values. The impact of this error on the selection of cut-off values is minor.

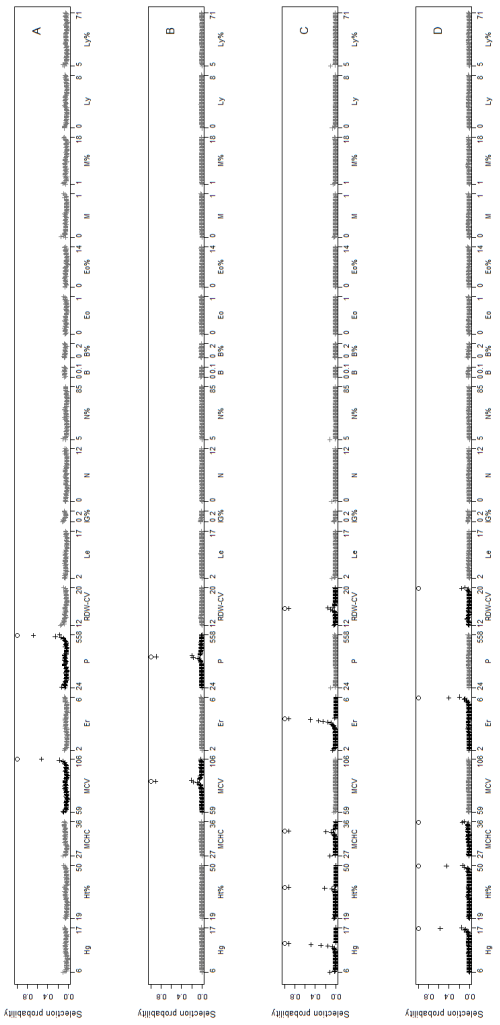

Figure 1: Selection probability for the 4 simulated scenarios. Values are mean over 100 simulation. Black and gray crosses indicate estimated values of relevant and irrelevant predictors, respectively. Circles indicate target values.

4. Conclusion

No multivariate analysis studies, even based on classical statistical methods, have been used to adjust the consensus cut-off values. In general, studies in the literature propose new cut-off values and compare them to the consensus ones using some criteria such as recall. However, they do not describe the process of deducing the new (more efficient) cut-off values. Likely, this process is based on acquired expertise which is not reproducible. Our procedure represents a good trade off between traditional interpretable models (Vellido, 2019) and powerful machine learning methods for clinical laboratories. It is able to detect true cut-off values with high probability, provided that enough data are available in their neighbourhood, while comparable in predictive performance to deep learning (results showed in Avalos et al. (2020)).

Our tool has a straightforward practical application from an open source code available on request from the corresponding author (shortly on https://github.com/mavalosf).

Acknowledgments

The ideas of this work were born from informal discussions between Marta Avalos and Marcela Henriquez during their research visit in Canberra, Australia (CSIRO's Data61 and ANU John Curtin School of Medical Research, respectively) in 2017. Ideas took shape during the participation of Marta Avalos to the "Big data: the information revolution in biomedical research" symposium organised by Marcela Henriquez and her colleagues from the Pontificia Universidad Catolica de Chile in December 2018.

Marta Avalos kindly thanks Cheng Soon Ong (CSIRO's Data61, Canberra) and Aditya Krishna Menon (formerly, CSIRO's Data61, Canberra; currently, Google, New York) for sharing their knowledge on imbalanced class learning. Marcela Henriquez is indebted to Teresa Quiroga from the Department of Clinical Laboratories, School of Medicine, Pontificia Universidad Catolica de Chile, Santiago, Chile for contributions regarding the CBC data and for advice.

The calculations were carried out on the CURTA cluster of Mesocenter for Intensive Calculation in the Aquitaine French region.

References

U Amato, A Antoniadis, and I De Feis. Additive model selection. *Stat Methods Appl*, 25:519–564, 2016.

M. Avalos, H. Touchais, and M. Henriquez-Henriquez. Optimising criteria for manual smear review following automated blood count analysis: A machine learning approach. In *Proceedings of the 10th WICT*, Advances in Intelligent Systems and Computing, Springer, 2020. To appear.

P W Barnes, S L McFadden, S J Machin, and E Simson. The international consensus group for hematology review: suggested criteria for action following automated CBC and WBC differential analysis. *Lab Hematol.*, 11:83–90, 2005.

I Barrio, I Arostegui, J M Quintana, and I C Group. Use of generalised additive models to categorise continuous variables in clinical prediction. *BMC Med Res Methodol*, 13(83), 2013.

A Chouldechova and T Hastie. Generalized additive model selection. Technical report, arXiv:506.03850v2, 2015. URL https://arxiv.org/abs/1506.03850.

D V Clé. Blood film in the era of streaming cells. *Rev Bras Hematol Hemoter*, 39(4): 295–296, 2017.

J P Dmochowski, P Sajda, and L C Parra. Maximum likelihood in cost-sensitive learning: Model specification, approximations, and upper bounds. *JMLR*, 11:3313–3332, 2010.

H He and E A Garcia. Learning from imbalanced data. *IEEE Trans on Knowl and Data Eng*, 21(9):1263–1284, 2009.

B Liquet and J Riou. CPMCGLM: an R package for p-value adjustment when looking for an optimal transformation of a single explanatory variable in generalized linear models. *BMC Med Res Methodol*, 19 (1):79, 2019.

Y Lou, J Bien, R Caruana, and Gehrke J. Sparse partially linear additive models. *J Comput Graph Stat.*, 25(4):1126–1140, 2016.

N Meinshausen and P Bühlmann. Stability selection. *J Roy Statist Soc Ser B*, 72:417–473, 2010.

A Petersen and D Witten. Data-adaptive additive modeling. *Stat Med*, 38(4):583–600, 2019.

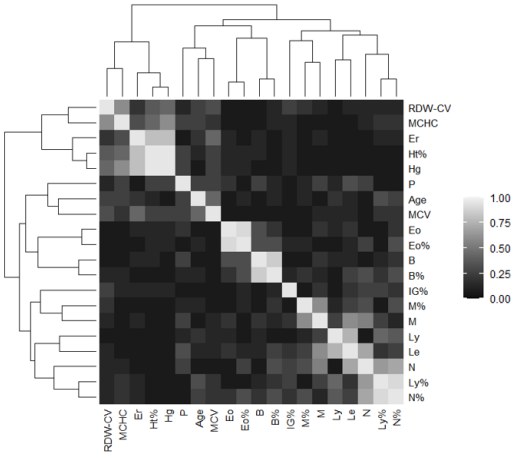

Figure 2: Heatmap of absolute value correlations of continuous predictors in the real CBC data.

R Tibshirani. Regression shrinkage and selection via the lasso. *J Roy Statist Soc Ser B*, 58:267–288, 1996.

A Vellido. The importance of interpretability and visualization in machine learning for applications in medicine and health care. *Neural Comput Appl*, pages 1–15, 02 2019.

Appendix A. Appendix

Supplementary figures are presented in the appendix.

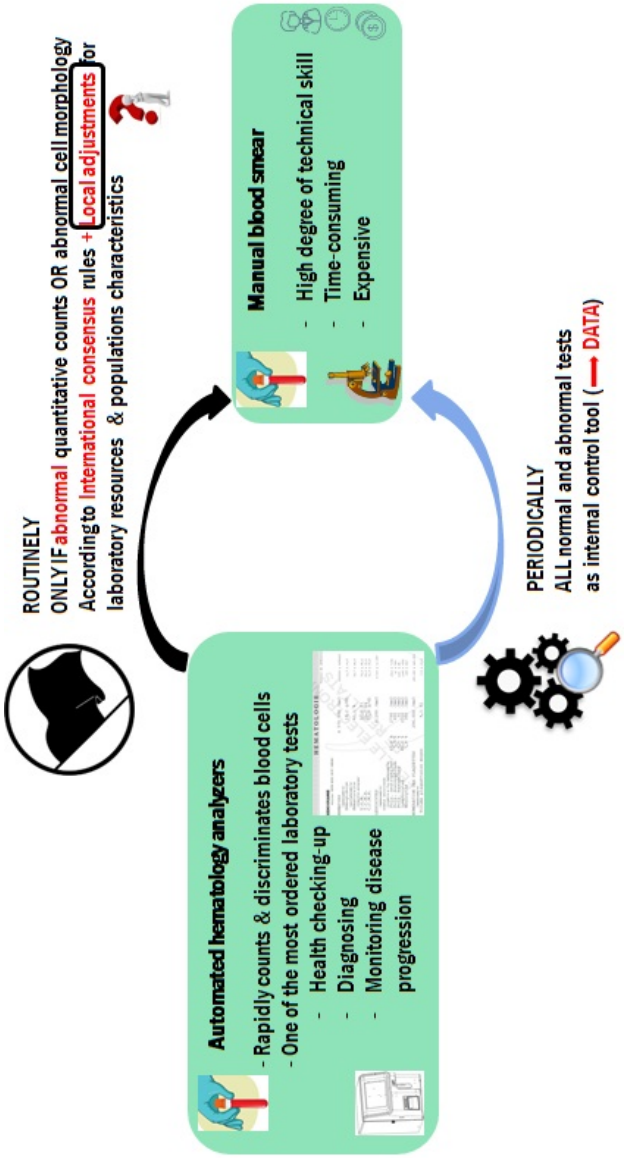

Figure 3: Scheme of the clinical laboratory procedures.