dHealth 2023
B. Pfeifer et al. (Eds.)

doi:10.3233/SHTI230008

Improving Blood Donor Care in a Collection Center Through Advanced Data Exploitation

Alice V. BOTTINELLI[a] , Silvia POZZI[a],
Alessia ZANNI[a] and Ettore LANZARONE[a,1]

[a] *University of Bergamo, Dalmine (Bg), Italy*

Abstract. Background: Blood collection centers can take advantage of the huge amount of data collected on donors over the years to predict and detect early the onset of several diseases, However, dedicated tools are needed to carry out these analyses. Objectives: This work develops a tool that combines available data with predictive tools to provide alerts to physicians and enable them to effectively visualize the history of critical donors in terms of the parameters that led to the alert. Methods: The developed tool consists of data exchanging functions, interfaces to raise alerts and visualize donor history, and predictive algorithms. It was designed to be simple, modular and flexible. Results: A prototype was applied to the Milan department of the *Associazione Volontari Italiani Sangue*, and was deemed suitable for prevention and early diagnosis objectives by the physicians of the center. The included Machine Learning predictive algorithms provided good estimates for the variables considered in the prototype. Conclusion: Prevention and early diagnosis activities in blood collection centers can be effectively supported by properly using and displaying donor clinical data through a dedicated software tool.

Keywords. Blood Donation, Early Diagnosis, Data Management, Machine Learning

1. Introduction

In Western countries, blood is usually collected from healthy individuals who donate it voluntarily and free of charge [1]. Typically, a donor goes to a blood collection center where he/she is examined by a physician and, if eligible, can donate [2]. At the same time as the donation, a blood sample is taken to the check donor's health condition and the usability of the blood unit. If the tests show no problems, the unit can be used and is sent to the next echelon of the blood supply chain [3,4], while the donor can return for the next donation after at least a given period established by law. Otherwise, the donor is suspended temporarily or permanently.

Together with the donation activity, the mission of some collection centers includes prevention activities and early diagnosis aimed at donors. More specifically, blood collection centers can take advantage of the huge amount of data collected on donors over the years to predict and detect early the onset of various diseases. At the same time,

[1] Corresponding Author: Ettore Lanzarone, University of Bergamo, Dalmine (Bg), Italy, E-Mail: ettore.lanzarone@unibg.it

the physicians of the collection center can use this information to give useful indications to donors, to correct their lifestyle and avoid the onset of pathologies. It is worth noting that, unlike other healthcare facilities, blood collection centers periodically collect data from healthy subjects for several years even before some of them develop a disease, i.e., a wealth of information on which to implement effective predictive tools.

However, these analyses require time and effort, as physicians should analyze a large amount of data and the available predictions about the onset of pathologies in order to have a complete overview of the donor's condition. But physicians, whose primary task is to carry out the collection of blood units, cannot regularly conduct these analyses except in cases they consider particularly serious.

In the literature, most of the decision support systems for blood donation refer to management aspects, e.g., organization and forecasting of next donations [5]. Only a few contributions concern donor health conditions and the use of visualization platforms and predictive tools to monitor donor status [6]. Moreover, other tools are available for the management of donated blood units [7].

In this light, a support tool that automatically processes the data, integrates it with predictive tools, and provides alerts to physicians to highlight cases that require their attention would be an effective support. Moreover, since physicians usually record the outcome of the pre-donation screening on a software, the integration of this support tool within the management software already in use at the collection center would guarantee its usability, as the alerts would be generated by the management software itself without the need to open other platforms.

The aim of our work was therefore to develop a tool that combines available data with predictive tools in order to profile donors and provide alerts on the onset of different diseases. Also, in critical cases, the tool allows the physician to effectively visualize the donor history in terms of the parameters that led to the alert, without sifting through the complete donor medical record.

The tool was developed in collaboration with the Milan department of the *Associazione Volontari Italiani Sangue* (AVIS), hereinafter referred to as AVIS Milan [8]. It is one of the largest blood collection centers in Italy, which can be considered as a general case of center in terms of management [9,10] and carries out several prevention and early diagnosis activities. The prototype, designed in collaboration with and validated by the AVIS Milan staff, was developed to be integrated with its management software and to use the data collected in its databases.

2. Methods

The tool is made up of three main parts, i.e., the functions for exchanging data with the AVIS Milan databases, the interfaces for raising alerts and displaying the history and predictions of the parameters that led to the alert, and the predictive algorithms.

It was designed to be simple, modular and flexible, and to be easily expanded in the future. Indeed, the architecture for analyzing a set of parameters and providing alerts for a condition or disease is common for all analyses performed, while the parametrization is different based on the specific analysis. Furthermore, the parameter values for the different analyses are stored in one of the AVIS Milan databases to allow for easy recalibration when required. This structure also allows easy traceability of calculations and expansion of the tool as new analyses are needed.

To pursue integration with the existing management software, the tool was implemented in Microsoft Visual Basic and data exchange was included through SQL queries. Instead, an R script was used to make predictions, called within the tool and ran in the background.

2.1. Data exchange

AVIS Milan stores its data in two different databases. The first database, according to the rules of the regional health system, collects all the information related to the donated units, to ensure their traceability and the exchange of information with other health facilities. Its structure is imposed by the regional health system and it is not possible to add other fields. The second database, created by AVIS Milan, allows storing all additional data regarding donor profiles and specific health information. This is necessary to carry on the activities of AVIS Milan concerning prevention and early diagnosis.

The information stored in these databases for each donor is of two types. A first group of data, which can be defined as static, characterizes the donor once and for all, e.g., sex, date of birth and other personal information. A second group, which can be defined as dynamic, is measured repeatedly each time a visit or donation is made and includes, e.g., the results of blood tests.

Moreover, for the functioning of the proposed tool, additional tables were added to the second database of AVIS Milan:

- *Parameters*: this table includes the list of tests and related alarms to be generated for each donor profile. For example, for each alarm, it includes information about which tests to consider (and therefore to perform) for any combination of donor sex, age range, and prior disease propensity. It also includes the frequency of execution (in terms of number of donations or months) of each of the tests listed.
- *Thresholds*: this table includes the physiological range for the values of the considered tests and the indices calculated from them. These ranges can also be differentiated according to the donor profile (e.g., sex and age range).
- *Alarms*: this table records all raised and non-raised alarms. Every time a donor is checked, the positive or negative onset of each alarm is stored in this table. This allows to use the alarms in every part of the tool, once computed, and to trace them for future checks.
- *Objectives*: this table includes and formalizes the goals that physicians set to promote healthy donor behaviors. For example, if a donor is at risk of developing a disease due to a bad habit, the physician can define a goal to achieve and the deadline for achieving it. Then, the physician and the donor can check if the goal is met within the deadline or reschedule it. Even if this is not closely related to the other parts of the tool, these objectives also contribute to the pursuit of prevention.

2.2. Interfaces

The main interface that summarizes the functionalities of the tool is shown in Figure 1. It can be opened either by choosing a donor from the list, using the already existing interfaces of the management software, or by clicking a button labeled *Details* when an alarm is raised. This interface provides the donor's information (only the ID, called CAI in AVIS Milan, and the sex are reported in the figure for privacy reasons) and the details of all variables analyzed.

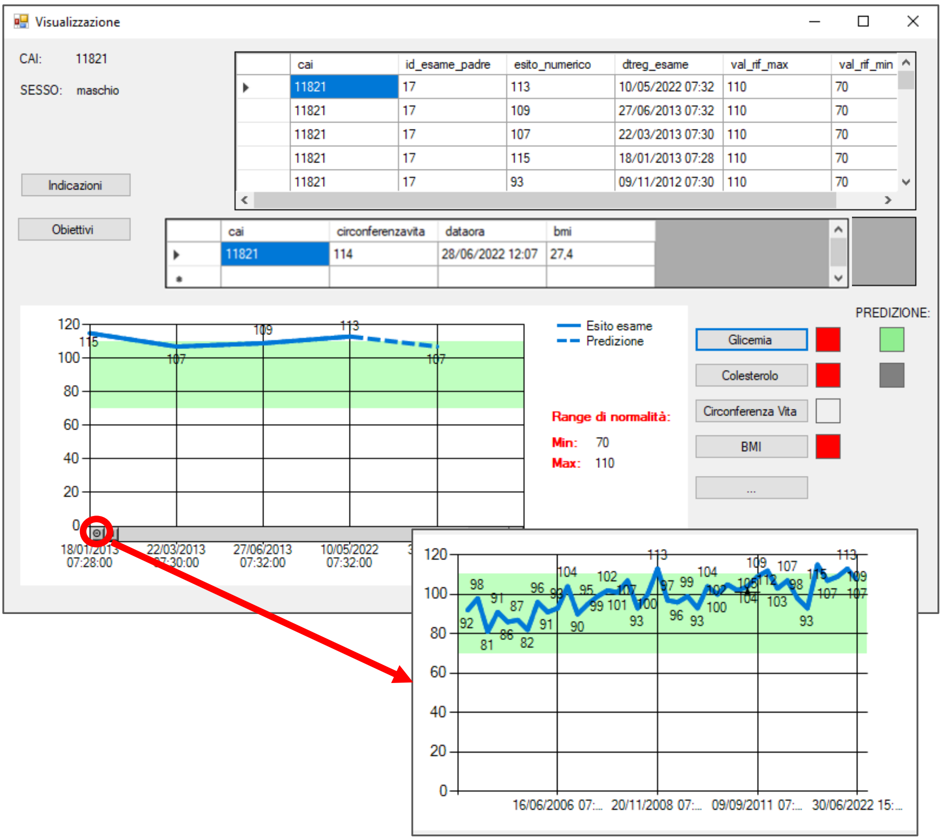

Figure 1. Interface with the main functionalities of the tool.

In the upper part of the window, two tables display the results of the last medical tests recorded for the donor. The first table includes the tests that are necessary for the regional healthcare system and stored in the first database, while the second table includes the additional data collected by AVIS Milan and stored in the second database. The division is maintained in the window to give physicians an idea of the added value of AVIS Milan prevention activities. For each test, the ID and the numerical result are reported together with the date and the physiological range for the donor under consideration; the ID can easily be replaced by the name if the physician prefers it.

The lower right part of the window displays the alerts for each test and combined metric through the color of the respective square: red means that the last measured value falls outside the physiological range, orange that the last measured value is in the physiological range but the penultimate and/or the antepenultimate are not; otherwise, the square remains transparent. An additional square on the right changes color according to the same logic, but based on predictions of future value and not actual measurements. An alert is triggered by the integrated management software when a measured or predicted value falls outside the physiological range.

By clicking on a button with the name of a test or metric, a graph appears on the left showing the time course of the measurements according to the dates together with the physiological range for the donor. The scroll bar allows to go back to previous results or focus on a specific period, while the small square button in the lower left corner allows to have a complete overview of the time course. Finally, the dashed line refers to the predicted values that are appended to the trend shown. The last button allows defining which tests or metrics are displayed in the form.

Another secondary interface displays the details of the objectives defined for each donor, together with their deadline and achievement status.

2.3. Predictive algorithms

We considered a set of Machine Learning (ML) approaches to predict the next values of a test or metric. In particular, following the procedure recently exploited in [11] for a different application context, we applied several approaches and chose the one with lowest 10-fold cross-validation error when applied to AVIS Milan historical data.

We assumed Markovianity in all approaches, i.e., the subsequent value is predicted only with the current value of the dynamic variables (those recorded at each donation) and fixed value of the static ones (those that remain constant for any given donor).

In particular, we tested the classical linear regression, several penalized regressions through the elastic net regularization (including ridge and lasso regressions), support vector machines with linear, polynomial or radial kernel, adaptive and polynomial adaptive splines, the random forest, and the Bayesian additive regression tree.

3. Results

We created our prototype including cholesterol and glycemia, which have a primary role in the metabolic syndrome and have been identified as highly important by the physicians of AVIS Milan. The prototype was integrated into the management software in use, and its technical feasibility was evaluated both in terms of data exchange and correctness of the interconnections.

Subsequently, the prototype and the interfaces were shown to the physicians working in AVIS Milan, i.e., the end users of the tool. They deemed the tool suitable for prevention and early diagnosis goals, and confirmed the easy access to information that allows them to quickly get information about patients at risk. Obviously, AVIS Milan physicians indicated that it is not sufficient to consider only cholesterol and glycemia for donor evaluation, but that the proposed structure can be effectively replicated for many other measurements and metrics without major modifications.

Concerning the ML approaches, a total of 37’863 and 45’885 observations were available for cholesterol and glycemia, respectively.

Random forest and Bayesian additive regression tree revealed too long computational times for training and were therefore excluded. Support vector machine with polynomial kernel, adaptive splines and polynomial adaptive splines performed worse than the other approaches in the considered 10-fold cross-validation. The other approaches showed similar values of normalized root mean squared error (NRMSE), although the support vector machine with radial kernel was the best approach for both predicted variables. Its NRMSE on the test set under the 10-fold cross-validation was equal to 11% for cholesterol and 14% for glycemia.

4. Discussion

The developed prototype proved effective in supporting the physicians of a blood collection center in their prevention and early diagnosis activities towards donors.

On the one hand, the richness and quantity of information contained in the databases of collection centers, e.g., AVIS Milan, allows for an in-depth analysis of donors' health conditions before they begin to suffer from any pathology. However, on the other hand, physicians cannot access all this information easily. Also, due to the limited time physicians visit a donor, it is not possible for them to analyze all aspects of the data and there is a risk of missing important indicators.

Therefore, a tool is needed that can support physicians in identifying situations at risk, so that the physicians focus only on them while carrying out the regular activity of collecting blood units. Our prototype, as confirmed by AVIS Milan, has this potential. Furthermore, being integrated into the management software already in use, it does not require physicians to carry out further duties, which would jeopardize its actual use in clinical practice

In the future, we will extend the prototype to include other measurements and metrics, and we will generalize it further so that it can be used in other blood collection centers as well.

References

[1] World Health Organization (WHO). *Blood donor selection: guidelines on assessing donor suitability for blood donation*. 2012.

[2] M. Doneda, S. Yalçindağ, I. Marques, E. Lanzarone. A discrete-event simulation model for analyzing and improving operations in a blood donation center. *Vox Sang* **116**(10) (2021), 1060–1075.

[3] W.P. Pierskalla. *Supply chain management of blood banks*. in: Operations research and health care: a handbook of methods and applications. Kluwer's International Series, 2004. pp. 103–145.

[4] J. Beliën, H. Forcé. Supply chain management of blood products: A literature review. *Eur J Oper Res* **217**(1) (2012), 1–16.

[5] S. Mahadevan, S. Poornima, K. Tripathi, M. Pushpalatha. A survey on machine learning algorithms for the blood donation supply chain. *J Phys Conf Ser* 1362 (2019), 012124.

[6] S. Bhingarkar. A comparison of machine learning techniques for categorization of blood donors having chronic hepatitis C infection. *Intelligent Data Communication Technologies and Internet of Things: Proceedings of ICICI 2021* (2022), 731-742.

[7] L. Tim Goodnough, S. Andrew Baker, N. Shah. How I use clinical decision support to improve red blood cell utilization. Transfusion 56(10) (2016), 2406-2411.

[8] AVIS Milan, www.avismi.it, last access: 28.1.2023.

[9] S. Baş Güre, G. Carello, E. Lanzarone, S. Yalçindağ. Unaddressed problems and research perspectives in scheduling blood collection from donors. *Prod Plan Control* **29**(1) (2018), 84–90.

[10] S. Baş, G. Carello, E. Lanzarone, S. Yalçindağ. An appointment scheduling framework to balance the production of blood bags from donation. *Eur J Oper Res* **265**(3) (2018) 1124–1143.

[11] M. Doneda, S. Poloni, M. Bozzetto, A. Remuzzi, E. Lanzarone. Surgical planning of arteriovenous fistulae in routine clinical practice: a machine learning predictive tool. *J Vasc Access* (2023), DOI 10.1177/11297298221147968.