Advancing Blood Supply Chain Prediction Based on a Novel Hybrid Machine Learning

Chaimae Mouncif, Mohamed Amine Ben RABIA, Adil Bellabdaoui
Team of Information Technology and Management (ITM), ENSIAS – Mohammed V University in Rabat, Morocco

***Abstract*—Blood supply chains constitute a critical yet often overlooked component of modern public health systems, as they coordinate donors, collection centers, hospitals, and patients. One of the major operational challenges lies in planning the deployment of mobile blood collection units under highly variable and uncertain spatio-temporal demand. In this context, this study proposes a novel hybrid machine learning framework for predicting donor return potential and supporting location and time selection decisions for mobile blood drives. The proposed approach combines Support Vector Regression (SVR) and Light Gradient Boosting Machine (LGBM) through a dynamic, context-aware weighting function designed to capture both temporal regularities and nonlinear spatial heterogeneity in donor behavior. The model is evaluated using real-world data collected from a blood collection center operating multiple mobile units. Experimental results demonstrate that the proposed hybrid framework consistently outperforms its individual components, achieving R^2 values of up to 83% for certain locations, together with low Mean Absolute Error (MAE) and Mean Squared Error (MSE). These results confirm the robustness and stability of the proposed approach. Beyond predictive performance, the model is intended to be integrated into a decision-support system to help managers optimize logistical resources and improve the strategic planning of mobile blood collection campaigns. This work contributes to the emerging field of data-driven blood supply chain optimization by introducing a spatio-temporal, hybrid predictive core specifically designed for operational decision support.**

***Keywords*—Blood supply chain; mobile blood collection units; spatio-temporal prediction; hybrid machine learning; decision support**

I. INTRODUCTION

Nowadays, blood supply chains are a fundamental but unobtrusive pillar of modern public health. They play a vital role in coordinating and optimizing the management of this important function between donors, blood centers, hospitals, and patients, who are the final link in the chain. For example, if a patient urgently requires a blood transfusion, the speed and efficiency of the blood management logistics chain will save his or her life and contribute to the safety and protection of human life [1].

These chains are not limited to collecting or transporting blood between collection centers and hospitals but are complex and rigorous functions involving upstream planning, disciplined execution, and reactivity coupled with real-time coordination [2]. The stakes are high because each component must function correctly, without fault or delay. The optimization of these chains using mathematical models from the world of decision analytics or management intelligence is therefore of the utmost importance.

Blood supply chain management can involve managing an orchestra in which each entity (donor, center, hospital, patient) must fulfill its responsibilities. It's a general order of execution that requires a high degree of coordination and is subject to strict rules to ensure optimization and flawless decision-making.

The variability of demand between different hospitals and regions is a major challenge. The blood supply is not a linear function, where we know exactly what is needed in a given place and at a specific time. On the contrary, needs are highly variable and can suddenly explode in one region without impacting others or in all delivery sites. In this context, it would be extremely useful to anticipate collections and needs, especially when we're trying to determine collection points precisely and avoid shortages.

Faced with the complexity of blood management, the digitization of processes is essential for effective and efficient management. At this stage, digital technologies such as big data analytics, the Internet of Things, RFID, computer vision and predictive technologies [3], [4] can clearly help to offer concrete solutions for increasing performance, responsiveness and optimization in management.

With the increasing digitization of the blood supply chain, voluminous data is generated at every stage of the chain's management process. From the donor's first step, through the transport of blood, its storage in collection centers, its distribution to hospitals and finally its use by doctors. In this context, using business analytics as an approach to analyzing trends and making the right decision at the right time is arousing growing interest within the scientific and academic community.

The use of business analytics, with an emphasis on its predictive aspect, enables detailed analysis of historical data collected in data warehouses, to anticipate future trends in terms of demand and need [5], [6]. However, despite the abundance of work using predictive analytics in the academic literature, its use for managing and optimizing decisions in blood supply chain management remains timid. Moreover, there is a lack of work on optimizing the location of mobile blood collection units on the basis of predictions and forecasts of blood donor concentration (Fig. 1).

Optimizing mobile blood drives is a major challenge for decision makers. In particular, when they want to program a campaign in a region requiring major mobilization of means and resources, but have no idea or indication of possible donors or the recidivist of previous donors in the region. By intelligently exploiting previously collected data using sophisticated machine learning methods, decision makers can base their decisions on reliable performance indicators, offering an all-round vision.

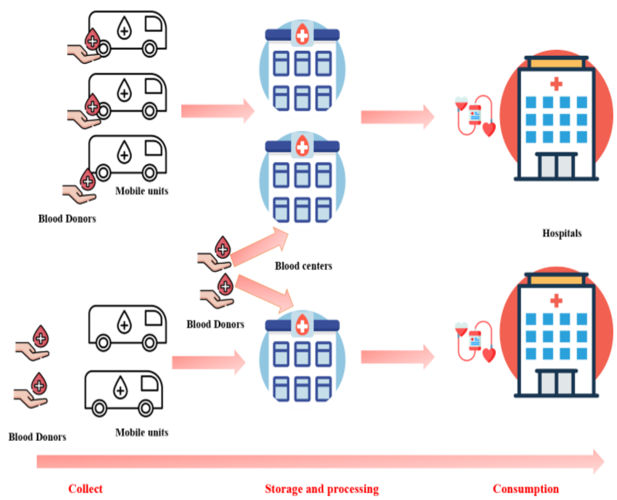

Fig. 1. Blood supply chain with mobile units.

With this in mind, our research involves collecting data on blood donors. These data come from a blood collection center in Morocco, with several mobile units. This study is very important, especially in the case of Morocco blood supply chain, where blood requirements are very high compared with the national stock of this material. There is a glaring discrepancy between blood collection and needs. Morocco needs more than 1,000 donations a day to be self-sufficient in labile blood products. With this study, we hope to provide a tool for organizing mobile blood collection campaigns, so as to guarantee an increase in the national stock to meet the needs of hospitals in a timely manner.

The aims of this study are to predict blood donor trends and streamline the movement of mobile units. To ensure a reliable and robust global prediction, our method attempts to integrate two machine learning models frequently used in the blood supply chain, SVR and LBGM, via a new function. Decision-makers will be able to make choices that reduce logistical expenditure and unnecessary travel, thanks to the integration of this newly-created framework into a dashboard.

Although numerous studies have investigated donor behavior prediction and others have proposed decision-support dashboards for blood supply chains, most existing works treat these two dimensions separately. Moreover, ensemble and hybrid models are often implemented using standard strategies such as stacking, boosting, or static weighted averaging. In contrast, this work addresses a specific and still underexplored problem: the spatio-temporal optimization of mobile blood collection unit deployment using a dynamic hybrid predictive mechanism. The proposed approach introduces a context-aware combination function that dynamically adjusts the contribution of SVR and LGBM according to the characteristics of the prediction task, rather than relying on a fixed or meta-learned ensemble structure.

We can list the major contributions of our work in the following points:

1) *Combined-output:* We aim to improve the performance of the support vector regression (SVR) model by combining its output with that of the light gradient machine (LGBM), two algorithms recognized for their effectiveness in supply chain analytics. The proposed approach is based on multiple inputs and incorporates a new combination function designed to enhance prediction accuracy.

2) *Donor return prediction:* The central objective is to anticipate the likelihood of a donor returning to make a future donation. Reliable predictions in this area enable blood collection organizations to design personalized loyalty strategies and interact more effectively with donors.

3) *Decision-making indicators:* Provide managers with decision-making indicators for planning the return of mobile blood collection units to a specific location based on the results generated by the predictive model.

The rest of the paper is organized as follows: Section II presents a review of related works. Section III describes methods used and presents the novel predicting function. Section IV explains data, results, and discusses the findings. Finally, we conclude the paper and propose future research directions in Section VI.

II. Literature Review

Predictive analytics has become a fundamental tool for prospecting potential donor pools and characterizing the factors determining the decision to donate blood or not [7], [8]. This study consolidates and critically evaluates original research from India, Saudi Arabia, Iran, Italy, Thailand, and East Asia, including methodological designs and theoretical predictive models of blood donation behavior. It also highlights the comparative performance of the models, general shortcomings, and limitations.

The main objectives of blood donation prediction research are to understand the prediction of individual donor behavior and the prediction of blood supply and demand as a whole [9]. Various methods are used, all adapted to the complexity of these tasks and presenting different advantages. Relevant work has used Bayesian methods, as shown in [10], analyzing donation intensity as a function of random individual frailties and covariates (demographic, health, and habit). Based on a large amount of data concerning 5,937 regular donors in Milan, the Bayesian method was used to identify frequent donors and predict future donations. To classify individual donors, machine learning approaches are common techniques such as K-nearest neighbors (KNN), Naive Bayes and decision trees [11]. [12] used logistic regression and achieved good results with a prediction score of around 69%, processing unbalanced data using Synthetic Minority Oversampling Technique (SMOTE). [13] use artificial intelligence and machine learning techniques to classify and determine the factors influencing organ donation based on blood data. Time-Series techniques are well suited for forecast at a large scale, especially in use. [14], for example, used backpropagation artificial neural networks for monthly forecasting and reached MSE ranging between 0.1407 and 0.4507. similarly, [15], that harnessed an ensemble of models, such as XGBoost, LGBM, and CATBoost that combined disparate data with up to an R^2 value of 0.8497.

Comparison of these varied methods shows that Bayesian models provide useful interpretability and temporal flexibility. Complex patterns can be learned well by machine learning classifiers, in contrast with the importance of predictors emphasized by regression. Time-series techniques, furthermore, are well-suited for measuring gross trends. Finally, the ensemble models work well, especially for large datasets. This methodological diversity highlights the complementarity of these instruments, each bringing furtherance to our ability to forecast blood donations.

Several potential predictors of blood donation behavior have been identified in previous studies, although level of agreement between studies is less consistent. Demographic variables such as male sex, younger individuals and higher levels of body weight are always found to be associated with higher frequencies of donations with some studies suggesting that women tend to donate more intensively [16], [17]. Specifically, Behavioral history is one of the strongest predictors of future engagement, in particular recency of donation and frequency of donations, where shorter periods between donations have been linked to sustained involvement. Health and lifestyle characteristics such as nonsmoking, nondrinking, physical activity, and greater hemoglobin levels all predict higher donation intensity. However, these are underexplored in machine learning studies [18]. Contextual factors, such as cultural practices or hospital need, also influence donation patterns, with variances stated across regions [12], [15].

III. Materials and Methods

Our methodology is based on the global architecture presented in Fig. 2, which results in a combined prediction, $y = f(y_1, y_2)$, aimed at estimating the probability that a former donor will or will not donate blood at a given geographical location. This research proposes a strategic decision support framework applied to the blood supply chain, integrating the training of machine learning algorithms and the optimization [27] of their performance using various hyperparameter tuning techniques. We have data on blood donors collected during mobile campaigns organized in the interior of Morocco. Each blood drive center has mobile units that plan campaigns in remote areas to collect blood. Through this study, we aim to provide decision-makers with indicators based on a new combination of machine learning algorithms. The aim is to optimize the programming of mobile campaigns and avoid ineffective planning that fails to achieve set objectives.

It is important to emphasize that the proposed hybridization strategy does not fall into the category of stacking, boosting, or classical ensemble learning. No meta-model is trained, and the models are not combined sequentially nor with fixed global weights. Instead, a dynamic combination function is introduced, allowing the relative influence of SVR and LGBM to vary according to the spatio-temporal context of the prediction. This design choice aims to better capture the dual nature of the problem: temporal regularities and nonlinear spatial heterogeneity.

L et $y_{l,t}$ denote the target variable representing the donor return intensity at location l during time period t. In this study, $y_{l,t}$ is defined as the number (or normalized rate) of donors who return to donate blood at location l within a predefined time window following a previous campaign. This variable is constructed by aggregating historical donation records at the spatio-temporal resolution relevant for mobile unit deployment planning. The prediction task therefore consists in estimating $\hat{y}_{l,t}$, the expected donor return potential for a candidate location–time pair (l, t).

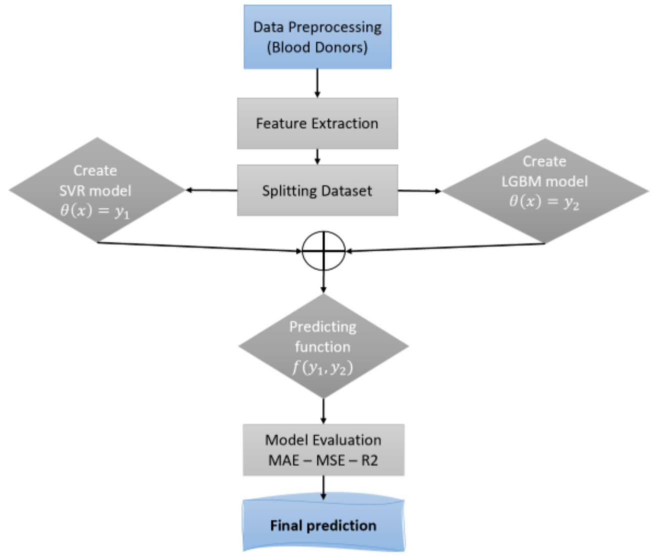

Fig. 2. Architecture of the proposed framework-based machine learning models.

A. Data Preprocessing

Indeed, acquisition errors, whether of human or technical origin, can alter the quality of the dataset and introduce biases when training models. These errors can take the form of incomplete information, missing or aberrant values, or even parasitic noise generated during data collection. It is therefore essential to implement a rigorous data preprocessing strategy to transform raw data into reliable, usable data, a prerequisite for obtaining high-performance machine learning models [19].

In this study, we focus on three key components of preprocessing: data cleaning, variable transformation and dimensional reduction. These steps aim to make the most of the analytical capabilities of the algorithms used, while ensuring the quality and consistency of the input data.

B. Features Selection

The feature selection principle shown in Fig. 3 represents a key method for reducing the dimensionality of a data set. Data sets generally contain many redundant and less informative variables, which can impact the relevance and performance of predictive models [20]. Moreover, it's important to highlight the most informative and relevant features for machine learning models [21].

Firstly, this technique reduces the complexity of the data, and therefore the complexity of the model. Small data sets, limited to the essentials, enable the model to learn quickly, thus reducing processing time and the need for more complex computing resources. Secondly, it significantly reduces noise by excluding from the study variables that are insignificant or have no real impact on the trend curve. This will contribute significantly to improving the quality of the dataset. Finally, based on its reduced data, the model will have a better chance of performing its trends and predicting phenomena to a degree close to reality [22].

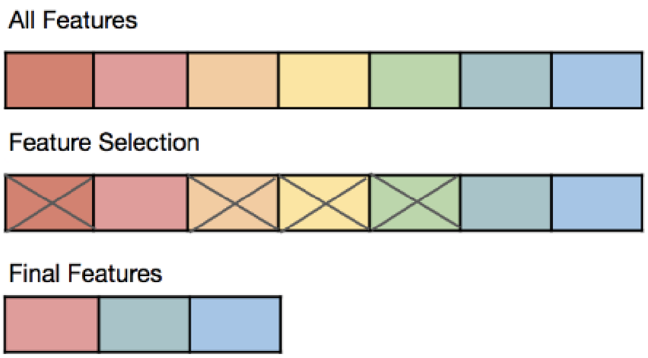

Fig. 3. Features selection steps.

C. Combined and Trained Model

Our main objective is to develop a high-performance prediction model based on the combination of two machine learning algorithms widely proven in the field of predictive blood supply chain: SVR and LGBM (Algorithm 1, Table I). This hybrid approach aims to capitalize on the complementary strengths of these two models:

TABLE I. SUMMARY OF VARIABLES AND EXPLANATION

Variable	Explanation
x_{i1}	Input value for SVR model
x_{i2}	Input value for LGBM model
y_{i1}	Predicted value for SVR model
y_{i2}	Predicted value for LGBM model
$\sigma\%$	Percentage of trained data
$1-\sigma\%$	Percentage of tested data
$e(y_{ij}, \hat{y}_{ij})$	Predicted error for model j
w_j	Weight for model j
W	*Normal Vector to the Hyperplan*
$f(y_1, y_2)$	Combined prediction function
ε	Error margin
b	Intercept of Hyperplane from Origin
C, ζ	Hyperparameter

SVR, recognized for its ability to model non-linear relationships with good generalization, and LGBM, appreciated for its efficiency, speed of learning and robustness in the face of noisy or incomplete data [23], [24].

By integrating a series of relevant attributes (inputs) related to blood donors and the logistical context, our system merges the predictions of the two models through a weighted aggregation function, to produce a single final prediction, more reliable and robust than those obtained individually. The combined prediction aims to improve the accuracy of estimates of blood donor behavior, while enhancing decision-making in proactive supply chain management.

Algorithm 1: Combined machine learning prediction algorithm

1: **Cleaning:** Removal of duplicates, management of missing values.

2: **Encoding:** Transformation of categorical variables.

3: **Standardization:** Scaling of numerical variables.

4: **Feature Selection:** Keep only the most relevant variables.

5: **Dataset:** we choose $\boldsymbol{\sigma}\%$ train data with $\mathbf{1}-\boldsymbol{\sigma}\%$ test data. We also add the condition that the test and validation data must be the most recent.

6: **Model training**: each model is trained separately. SVR model $\boldsymbol{\theta(x) = y_1}$ and LGBM model $\boldsymbol{\ddot{\theta}(x) = y_2}$

7: **For** i=1 to N

8: Predict the probability that a former donor $\boldsymbol{x}$ will donate blood in location $\boldsymbol{L}$ or not.

9: Calculate the prediction error per model between predicted $\boldsymbol{y_{i1}, y_{i2}}$ values and test values $\boldsymbol{\hat{y}_{i1}, \hat{y}_{i2}}$ $\boldsymbol{e(y_{i1}, \hat{y}_{i1}) = |y_{i1} - \hat{y}_{i1}|}$ ***and*** $\boldsymbol{e(y_{i2}, \hat{y}_{i2}) = |y_{i2} - \hat{y}_{i2}|}$

10: **End For**

11: **Generate the weight $\boldsymbol{w_1, w_2}$** for SVR and LGBM model based on their prediction error such as:

$$\boldsymbol{w_1} = \frac{\sum_{j=1}^{n} e(y_{j1}, \hat{y}_{j1})}{\sum_{j=1}^{n} e(y_{j1}, \hat{y}_{j1}) + \sum_{j=1}^{n} e(y_{j2}, \hat{y}_{j2})} \quad \text{and}$$

$$\boldsymbol{w_2} = \frac{\sum_{j=1}^{n} e(y_{j2}, \hat{y}_{j2})}{\sum_{j=1}^{n} e(y_{j1}, \hat{y}_{j1}) + \sum_{j=1}^{n} e(y_{j2}, \hat{y}_{j2})}$$

12: Calculate combined prediction $\boldsymbol{f(y_1, y_2) = w_1 y_1 + w_2 y_2}$

13: Calculate **MSE, MAE, R2**

14: **Validate** the combined prediction

15: **Generate decision KPI** based prediction

To guarantee the reliability and relevance of our model, we used a real dataset comprising approximately 15,000 observations, each described by 20 initial explanatory variables. To simulate a realistic context and preserve temporal consistency, we ordered the data chronologically. Thus, the 20% most recent observations were reserved for the test and validation phases. This approach enables us to assess the model's ability to generalize to future data, which is crucial for predictive applications, such as blood supply chain management.

SVR is an extension of the Support Vector Machine (SVM) algorithm, specifically adapted to regression tasks. It aims to find a function that minimizes the prediction error while maintaining a certain tolerance (called margin ε) to the training data. In other words, SVR constructs a "tube" around the regression function, in which data points are considered well predicted if they remain within this margin [25].

The SVR process is based on a series of steps aimed at finding a regression function capable of predicting accurately while tolerating a certain margin of error. The fundamental idea is to penalize predictions only when they deviate beyond a predefined margin ε. This makes it possible to model data more flexibly and robustly, especially when it is noisy. Input data is transformed using a kernel function, which projects the data into a higher-dimensional space. This transformation makes it possible to use a linear model in this transformed space, even if the original data show non-linear relationships. This process is essential if SVR is to adapt to complex data structures.

The model defines a “tolerance tube” around the regression function, delimited by the margin ε. Data points that fall within this tube are considered sufficiently close to the prediction and are therefore not penalized. On the other hand, points outside the tube are subject to a penalty proportional to their distance from the margin, thus controlling the impact of large errors.

The core of the process consists in minimizing a loss function, which takes two elements into account:

- On the one hand, the prediction error for points outside the margin ε,
- Secondly, the complexity of the model, measured by the width of the tube or the norm of the coefficients in the case of a linear kernel. This dual objective guarantees a good compromise between predictive performance and generalization capability.

The choice of kernel is a decisive step in the proper functioning of the SVR. Depending on the nature of the data, we can opt for a linear kernel (for simple relationships), polynomial kernel (for more complex relationships), or RBF (Radial Basis Function, or Gaussian kernel), which is the most commonly used, as it can model a wide variety of non-linear shapes.

Mathematical model of SVR:

$$argmin_{(w,b)} = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$subject\ to: y_i - (w^t x_i + b) \le \zeta$$

$$(w^t x_i + b) - y_i \le \zeta$$

$$for\ all\ \zeta \ge 0$$

LGBM is an open-source machine learning framework based on decision trees. It implements an ensemble technique called gradient boosting, which involves combining several weak models (typically shallow trees) to form a powerful predictive model.

BOOSTING is an esemplastic technique that consists of aggregating classifiers (models) developed sequentially on a training sample, with the individual weights corrected as they are learned. Classifiers are weighted according to their performance (Fig. 4).

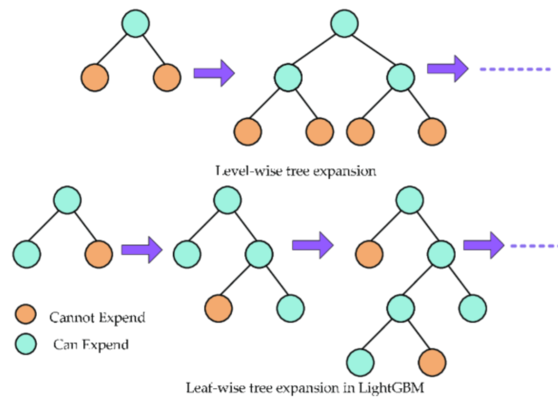

Fig. 4. Diagram of leaf-wise tree expansion.

D. Model Evaluation

The search for a model whose empirical error is minimal over a given set of observations. However, minimizing this empirical error does not guarantee minimizing the model error over the entire data space. Indeed, in an overlearning situation, the model error will be underestimated [26]. However, it is this error - or, in other words, our ability to make predictions about things that are not known - that interests us. This chapter shows how we can set up an experimental framework that allows us to evaluate a model while avoiding the bias of overlearning. With this in mind, we will make a distinction between evaluating a model, which consists in determining its performance over the entire data space, and selecting it, which consists in choosing the best model from among several.

In the case of a regression problem, the number of errors is not an appropriate criterion for assessing performance. On the one hand, because of numerical inaccuracies, it's tricky to tell from a true-value prediction whether it's correct or not. We attempt to evaluate our combination of machine learning models based on three metrics widely used in the field of artificial intelligence: MSE, MAE, R2.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

$$\mathrm{R2} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$$where\ \bar{y}\ is\ the\ average\ of\ observations$$

IV. Results

A. Data

The data used in our study are collected from 4 Moroccan blood collection centers with mobile blood collection units. Each center schedules its own campaign, with mobile units traveling to different areas to offer a service close to blood donors.

The problem is that the managers at the centers have no visibility of the points where former donors are most likely to return. The centers have a large amount of data covering the period 2019-2024, with almost 15,000 registrations. The data used consists of columns on donor ID, donor gender, donor age, date and time of each donation, location (x, y) of the mobile unit at the time of each donation, frequency of donation, time since last donation, whether the donor donated in the last campaign in October 2024, total quantity of blood donated.

Data cleaning was carried out to prepare reliable data for the prediction model. Empty data were filled in with 0 and incomplete data were completed in consultation with the center responsible. In addition, profiling work was carried out, especially on the descriptive analytics part, during which dashboards and descriptive statistics, using business intelligence tools such as Power BI, were delivered to each collection center's decision support center, accessible via a web portal (Fig. 5).

Fig. 5. Descriptive analytics of blood donation.

B. Results

For the results, we plan to make predictions by zone to determine where the mobile blood collection unit will be able to move. The analysis will also be carried out by five-day time windows. In order not to overload the article with too many graphs, we have chosen to visualize four zones, which we assume to be representative of the cases a decision-maker would wish to analyze.

We implement our models to compare the combined prediction with predictions using SVR and LGBM alone. We're using a powerful machine with over 64G RAM, 32 CPUs and 1 H100 GPU, which has enabled us to cut learning times and reduce processing times.

The results show that our framework for combining machine learning models performs better than the models alone. Fig. 6 shows that for a time window of 5 days for the locality 1 located at coordinates (Long., Lat.): (-6.8524783, 33.8548111), the combined prediction is highly relevant. With almost a slight discrepancy between actual and predicted values. On the other hand, the other models are more or less far from the actual values. Moreover, Fig. 7 shows that R2 scores very well, with a maximum of 83% for locality 1 and for the day of 03/10/2024.

TABLE II. FORECASTING PERFORMANCE OF MACHINE LEARNING METHODS BY LOCALITY

Models	Locality 1	Locality 2	Locality 3	Locality 4
Combined Model - MAE	33,36	15,2	17,01	29,12
Combined Model - MSE	25,12	39,12	14,12	23,26
Combined Model - R2	81,00%	77,15%	80,45%	69,21%
SVR - MAE	54,12	43,23	23,3	67,56
SVR - MSE	78,11	45,43	89,21	88,32
SVR - R2	55,32%	60,65%	68,43%	34,23%
LGBM - MAE	44,78	88,98	78,77	90,59
LGBM - MSE	70,34	91,234	89,221	114,24
LGBM - R2	65,43%	70,12%	88,19%	93,23%

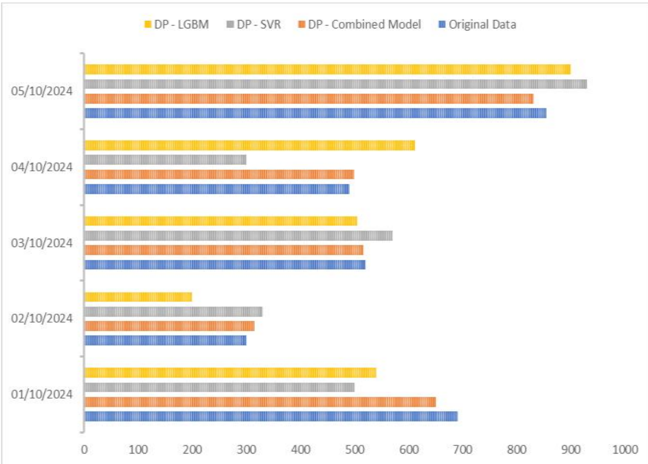

Fig. 6. Comparison of the predictions of the different models for the locality 1 located at coordinates (Long., Lat.): (-6.8524783, 33.8548111).

Fig. 7. Original data vs. Predicted data and R2 score for the locality 1 located at coordinates (Long., Lat.): (-6.8524783, 33.8548111).

We find almost the same result for locality 2 located at coordinates (Long., Lat): (-7.9432221, 31.3594816) as for locality 1, for a time period from 10/09/2024 to 14/09/2024. Fig. 8 also shows that the combined prediction best represents reality and reflects the actual data. Whereas SVR and LGBM performed alone for locality 2 and for the same time period do not show relevant results.

Fig. 9 also shows that a comparison with R2 shows that the coefficient displays a maximum percentage of 79% with values adjusted between actual data and blood donor prediction data.

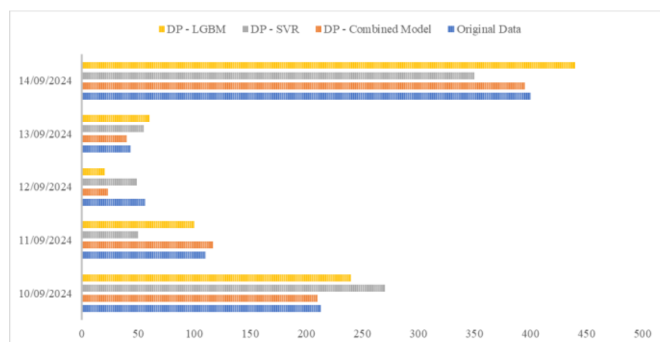

Fig. 8. Comparison of the predictions of the different models for the locality 2 located at coordinates (Long., Lat): (-7.9432221, 31.3594816).

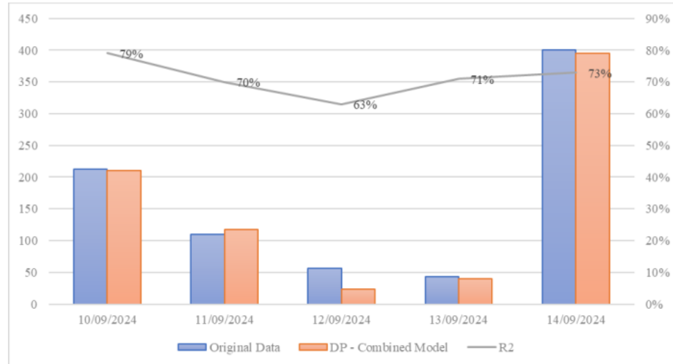

Fig. 9. Original data vs. Predicted data and R2 score for the locality 2 located at coordinates (Long., Lat.): (-7.9432221, 31.3594816).

- A more detailed comparison of the data by locality shows the effectiveness of combining the two models to produce a single prediction. In Table II, we have shown each model in relation to the three performance evaluation metrics MSE, MAE and R2 for 4 former localities where mobile units collected blood, and which we selected at random. The results show that the combined prediction always achieves significant and interesting scores compared with the predictions of the models alone.
- Table III and Fig. 10 also show that on the basis of a global prediction without taking blood collection locations into account, the combined prediction always has the best scores on the three performance evaluation metrics MSE, MAE and R2.

TABLE III. COMPARISON OF MODEL PERFORMANCE

Models	MAE	MSE	R2
Combined Model	33,36	32,22	81,16%
SVR	73,00	87,12	44,92%
LGBM	65,43	88,98	67,20%

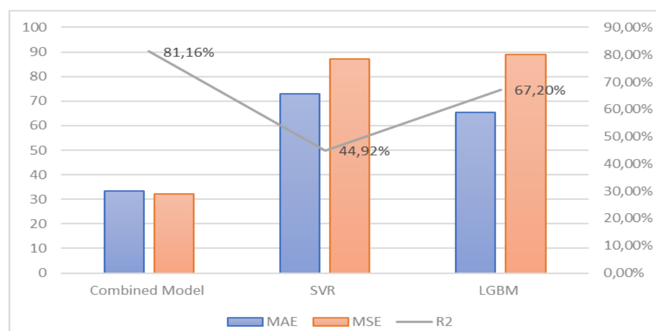

Fig. 10. Comparison of model performance.

V. DISCUSSION

When it comes to deciding on the location with the highest blood donor return rate, analytics is the ultimate solution to help decision-makers choose the best location for their mobile blood collection units. To this end, we have focused on prediction by machine learning models widely used in the blood supply chain, namely SVR and LGBM. We didn't just apply them as they are, but introduced a new concept of combination between the two models through a combined prediction function.

The strength of combining SVR and LGBM lies in the complexity of the resulting prediction. Predicting whether or not a former donor can give blood in future mobilizations. In other words, time and space. Gradient boosting has an excellent reputation in the data scientists' community, as it easily adapts to the complexity of inputs and outputs. For its part, SVR is also a good machine learning model that adapts to fluctuating data and outliers, enabling better results to be achieved when making predictions based on the time dimension.

The results we obtained from this combination, shown above in Table III, demonstrate the increasing performance of this combination technique on the weight of each model. At this stage, managers choosing either the campaign date or the campaign location can see performance indicators on whether or not a blood donor will return for the new blood drive. This will enable them to save resources in terms of personnel and means, by targeting only those localities with a higher recidivism rate. This predictive module based on the combination of SVR and LGBM can be integrated into a portal or management system for blood drives, giving them a futuristic dimension.

Although it is not possible to build a forecasting system for blood donor trends that perfectly anticipates changes in blood collection behavior, we use the results of the forecasting model to help decision-makers make judgments about the feasibility of the blood collection campaign. In this study, each forecasting model has a reference value. In the combined model fitting results, the maximum R2 value reached 81.16% for the whole dataset, and a maximum value of 83% for locality 1. In addition, the MAE and MSE showed values of 33.36 and 32.22 respectively, demonstrating a high level of prediction and robustness. The model showed a stable and excellent trend for this type of data, compared with the SVR and LGBM performed at each location.

VI. Conclusion and Recommendations

In summary, this study proposes a machine learning–based approach to identify locations with the highest blood donor return rates. We developed a hybrid prediction framework capable of capturing both the temporal and spatial dimensions of donor behavior by combining two widely used models in blood supply chain analytics: Support Vector Regression (SVR) and Light Gradient Boosting Machine (LGBM). This hybridization leverages LGBM's ability to model complex nonlinear relationships and SVR's robustness to data variability and outliers. The results obtained, particularly an R^2 of up to 83% for certain localities, together with low MAE (33.36) and MSE (32.22) values, demonstrate the relevance, robustness, and stability of the proposed model.

By accurately identifying regions with high donor return potential, the proposed approach helps decision-makers at blood collection centers optimize logistical resources and improve campaign planning. When integrated into a decision-support portal or management information system, this predictive framework can effectively guide strategic choices regarding the deployment of mobile blood collection units. Although the proposed tool cannot guarantee perfect forecasting accuracy, it constitutes a valuable medium- to long-term strategic lever for improving blood supply chain management.

Despite these encouraging results, several limitations should be acknowledged. First, the dataset is restricted to a single blood collection center in Morocco, which may limit the generalizability of the findings to other geographic or organizational contexts. Second, the temporal scope of the data may not fully capture long-term structural changes in donor behavior. Third, the proposed model relies on historical patterns and therefore cannot fully anticipate sudden behavioral shifts or epidemiological disruptions. Future work will aim to validate the approach on multi-center and multi-country datasets, integrate real-time and socio-demographic variables, and further refine the proposed dynamic weighting mechanism.

- Data Availability.
- Data will be made available on request.
- Conflict of interest.
- There is no interests to declare.

- Financial support and sponsorship.
- No sponsorship or funding supported this study.

References

[1] G. Dharmarathne, M. Bogahawaththa, U. Rathnayake, and D. P. P. Meddage, "Integrating explainable machine learning and user-centric model for diagnosing cardiovascular disease: A novel approach," Intell. Syst. with Appl., vol. 23, p. 200428, 2024.

[2] A. Ala, V. Simic, N. Bacanin, and E. B. Tirkolaee, "Blood supply chain network design with lateral freight: A robust possibilistic optimization model," Eng. Appl. Artif. Intell., vol. 133, 2024.

[3] R. Bhandawat, S. Casucci, B. Ramamurthy, and J. L. Walteros, "Cooperative Blood Inventory Ledger (CoBIL): A decentralized decision-making framework for improving blood product management," Comput. Ind. Eng., vol. 172, p. 108571, 2022.

[4] D. Kirk, C. Catal, and B. Tekinerdogan, "Predicting Plasma Vitamin C Using Machine Learning," Appl. Artif. Intell., vol. 36, no. 1, 2022.

[5] A. M. Ghouri, H. R. Khan, V. Mani, M. A. U. Haq, and A. B. Lopes de Sousa Jabbour, "An Artificial-Intelligence-Based omnichannel blood supply chain: A pathway for sustainable development," J. Bus. Res., vol. 164, 2023.

[6] M. Benaissa and S. Kaalda, "Genetic Algorithm: An Overview And View Through The Scope Of Data Science," Int. J. Sci. Appl. Prosper., vol. 1, no. 1, pp. 6–12, 2023.

[7] S.-S. M. Ajibade et al., "Evolution of machine learning applications in medical and healthcare analytics research: A bibliometric analysis," Intell. Syst. with Appl., vol. 24, p. 200441, 2024.

[8] Y. Harakay and M. E. Khaldi, "Augmented Business Intelligence: A Multi-Criteria Approach," Int. J. Sci. Appl. Prosper., vol. 3, no. 1, pp. 38–48, 2025.

[9] M. Chau, C. K. Lee, C. W. Chan, and E. Cheng, "Data analysis for healthcare: A case study in blood donation center analysis," 16th Am. Conf. Inf. Syst. 2010, AMCIS 2010, vol. 2, pp. 924–931, 2010.

[10] I. Epifani, E. Lanzarone, and A. Guglielmi, "Predicting donations and profiling donors in a blood collection center: a Bayesian approach," Flex. Serv. Manuf. J., no. 0123456789, 2023.

[11] K. Pabreja and A. Bhasin, "A Predictive Analytics Framework for Blood Donor Classification," Int. J. Big Data Anal. Healthc., vol. 6, no. 2, pp. 1–14, 2021.

[12] A. S. Alkahtani and M. Jilani, "Predicting return donor and analyzing blood donation time series using data mining techniques," Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 8, pp. 113–118, 2019.

[13] M. Amini, A. Bagheri, and D. Delen, “An explanatory analytics model for identifying factors indicative of long- versus short-term survival after lung transplantation,” Decis. Anal. J., vol. 3, p. 100058, 2022.

[14] A. A. J. S. Frengky Tedy, Patrisius Batarius, Ign. Pricher A. N. Samane, “Artificial Neural Network Based Prediction Model Back Propagation on Blood Demand and Blood Supply,” RESTI, vol. 7, no. 6, pp. 1403–1411, 2023.

[15] H. J. Kwon, S. Park, Y. H. Park, S. M. Baik, and D. J. Park, “Development of blood demand prediction model using artificial intelligence based on national public big data,” Digit. Heal., vol. 10, 2024.

[16] S. Kheiri and Z. Alibeigi, “An analysis of first-time blood donors return behaviour using regression models,” Transfus. Med., vol. 25, no. 4, pp. 243–248, 2015.

[17] A. Shrivastava, M. Chakkaravarthy, and M. A. Shah, “A new machine learning method for predicting systolic and diastolic blood pressure using clinical characteristics,” Healthc. Anal., vol. 4, p. 100219, 2023.

[18] S. Rekabi, Z. Sazvar, and R. Tavakkoli-Moghaddam, “A data-driven approach to optimize a blood supply chain within the Industry 5.0 framework: A stochastic optimization model,” Expert Syst. Appl., vol. 289, p. 127960, 2025.

[19] F. do R. Santos and R. Choren, “Data preprocessing for machine learning based code smell detection: A systematic literature review,” Inf. Softw. Technol., vol. 184, p. 107752, 2025.

[20] S. Yan, J. Qian, Y. Yu, and Y. Ni, “A feature selection method driven by fuzzy implication granularity,” Eng. Appl. Artif. Intell., vol. 158, p. 111298, 2025.

[21] L. K. Andersen and B. J. Reading, “A supervised machine learning workflow for the reduction of highly dimensional biological data,” Artif. Intell. Life Sci., vol. 5, p. 100090, 2024.

[22] J. El-Taraboulsi, C. P. Cabrera, C. Roney, and N. Aung, “Deep neural network architectures for cardiac image segmentation,” Artif. Intell. Life Sci., vol. 4, p. 100083, 2023.

[23] S. Alsulamy, “Predicting construction delay risks in Saudi Arabian projects: A comparative analysis of CatBoost, XGBoost, and LGBM,” Expert Syst. Appl., vol. 268, p. 126268, 2025.

[24] L. Guo, W. Fang, Q. Zhao, and X. Wang, “The hybrid PROPHET-SVR approach for forecasting product time series demand with seasonality,” Comput. Ind. Eng., vol. 161, p. 107598, 2021.

[25] H.-S. Yan and Z.-G. Shang, “Method for Product Design Time Forecasting Based on Support Vector Regression with Probabilistic Constraints,” Appl. Artif. Intell., vol. 29, no. 3, pp. 297–312, 2015.

[26] S. Hur et al., “Development, validation, and usability evaluation of machine learning algorithms for predicting personalized red blood cell demand among thoracic surgery patients,” Int. J. Med. Inform., vol. 191, p. 105543, 2024.

[27] D. H. Wolpert and W. G. Macready, “No free lunch theorems for optimization,” IEEE Trans. Evol. Comput., vol. 1, no. 1, pp. 67–82, 1997.