Explainable Machine Learning for Predicting Poor Mobilizers in Allogeneic Donors for Hematopoietic Stem Cell Transplantation

Dung Manh Nguyen*, Duy Nguyen*, Nghia Thi Mai†, Kou Yamada*
* Graduate School of Science and Technology, Gunma University, Gunma, Japan
{t241b601, t232b604, yamada}@gunma-u.ac.jp
† Posts and Telecommunications Institute of Technology, Hanoi, Vietnam, nghiamt@ptit.edu.vn

***Abstract*—Hematopoietic Stem Cell Transplantation (HSCT) requires sufficient mobilization of stem cells from healthy allogeneic donors. However, some donors, termed poor mobilizers, fail to yield adequate cells despite mobilization, leading to delays and adverse outcomes. To address this challenge, we propose an explainable machine learning framework that classifies donors into poor and good mobilizers using clinical features. The method integrates Recursive Feature Elimination and the XGBoost classifier into a lightweight pipeline suitable for deployment in multi-center, network-based clinical decision-support systems. The proposed model achieves 96.1% accuracy, 98.7% AUC, and 98.1% recall for poor mobilizers. To enhance interpretability, we employ SHAP to quantify the contributions of key clinical factors such as Platelet count, Age, and MCV. These explanations confirm established predictors while suggesting exploratory insights, supporting the potential of explainable Machine Learning (ML) to improve donor screening in HSCT.**

***Index Terms*—Explainable Machine Learning, SHAP, Hematopoietic Stem Cell Mobilization.**

I. Introduction

Hematopoietic stem cell transplantation (HSCT) is a curative therapy for hematologic malignancies and severe blood disorders. A critical step for successful transplantation is the mobilization of hematopoietic stem cells (HSCs) from the bone marrow into peripheral blood, usually induced with granulocyte colony-stimulating factor (G-CSF). Adequate mobilization is required to collect sufficient CD34+ cells for engraftment. However, a subset of donors, known as poor mobilizers, fail to achieve the required yield despite standard regimens, resulting in repeated procedures, increased costs, and treatment delays [1]. Early identification of poor mobilizers is therefore essential for optimizing donor selection and transplant planning.

Conventional predictors such as donor age, body mass index (BMI), and blood indices are readily available but often lack accuracy and fail to capture nonlinear interactions [2]. In recent years, machine learning (ML) approaches have demonstrated significant potential in biomedical prediction tasks, including donor classification and transplant outcome forecasting. Nevertheless, most existing studies focus primarily on achieving high predictive accuracy while neglecting the interpretability of the models. The resulting "black-box" nature of ML remains a major barrier to clinical use, as physicians require transparent reasoning to trust predictions [3]. In addition to clinical relevance, predictive models for donor mobilization can support connected and distributed HSCT workflows, where donor evaluation often occurs across multiple hospitals and laboratories. Lightweight and explainable ML models are increasingly important in such networked clinical environments, as they reduce communication overhead, allow transparent model sharing, and enhance reliability within federated or cloud-edge medical systems. Therefore, an interpretable and feature-efficient ML framework may contribute not only to clinical decision-making but also to broader information networking infrastructures that enable real-time donor assessment across institutions.

To address both the clinical and system-level needs, we propose an explainable ML framework using the XGBoost classifier combined with SHapley Additive exPlanations (SHAP). This approach not only achieves high predictive accuracy for identifying poor mobilizers, but also provides transparent explanations of feature contributions, offering insights at both global and individual levels. To the best of our knowledge, this is the first study to apply SHAP for explaining donor mobilization outcomes in HSCT. Our approach not only confirms established predictors such as platelet count and age, which are well supported by clinical literature, but also highlights exploratory factors such as mean corpuscular volume (MCV) and albumin, which may offer novel biological insights into mobilization heterogeneity. By bridging predictive performance with interpretability, this study contributes both a practical screening tool for clinicians and a deeper understanding of the biological variability in HSC mobilization, thereby supporting more reliable and transparent decision-making in HSCT donor management.

II. Proposed Method

The proposed workflow consists of five steps, as illustrated in Fig.1. First, donor data mobilized with G-CSF are preprocessed to handle missing values and noise, and the Synthetic Minority Oversampling Technique (SMOTE) is applied to address the class imbalance between poor and good mobilizers. The dataset is subsequently divided into 80% for training and 20% for testing to enable model development and independent evaluation.

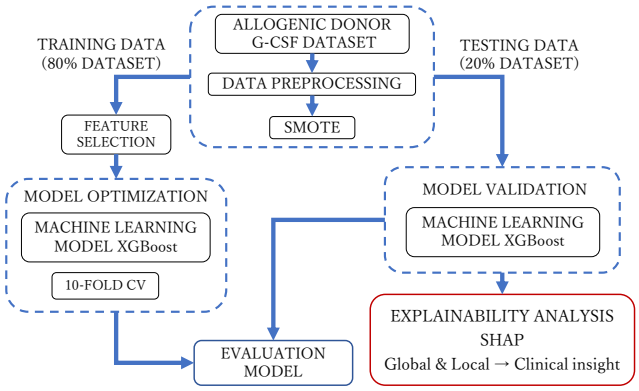

Fig. 1. Overview of the proposed method.

Second, feature selection is conducted using Recursive Feature Elimination (RFE) in order to remove irrelevant or redundant predictors and to retain a subset of clinically meaningful features. This step ensures that the model is trained on variables that contribute most to mobilization outcomes, reducing overfitting and improving interpretability.

Third, the XGBoost classifier is trained on the selected features. Hyperparameters are optimized through 10-fold cross-validation on the training set, providing a robust estimation of model performance and preventing overfitting. The optimized model is then evaluated on the held-out testing set and further validated on the entire dataset (excluding SMOTE-generated samples). Performance is assessed using commonly reported metrics in medical ML applications, including accuracy, AUC, and recall, with particular emphasis on recall for poor mobilizers due to its clinical importance.

Finally, explainable machine learning techniques are applied to enhance the transparency of the model. In this study, SHAP (SHapley Additive exPlanations) is employed to quantify the contribution of each feature to the model's predictions. SHAP provides both global explanations - ranking features by their overall impact on predictions across all donors- and local explanations that illustrate how specific features influence the outcome for individual donors. This dual perspective enables a deeper understanding of the relationship between clinical factors and mobilization outcomes, while also offering interpretable evidence to support decision-making in donor screening.

A. Data

The dataset consisted of 799 allogeneic donors, shared by Professor John F. DiPersio at Washington University [4]. Donor mobilization outcome is defined based on CD34+ cell counts after G-CSF administration: 171 donors with CD34 counts $< 40/\mu L$ are classified as poor or less-than-optimal mobilizers (class 0), while 628 donors with CD34 counts $\geq 40/\mu L$ are classified as good mobilizers (class 1). The input features include several clinically relevant variables such as Age, Sex, Body Mass Index (BMI), Platelet count, and other standard laboratory measures.

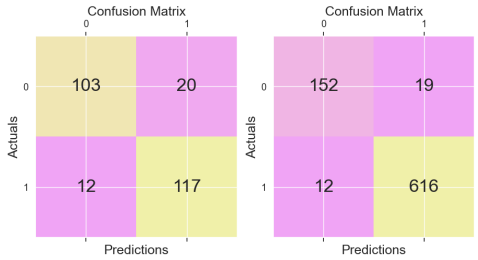

Fig. 2. Confusion matrices: (left) 20% testing set, (right) entire dataset. Class 0: Poor/less-than-optimal mobilization and Class 1: Good mobilization

TABLE I
PERFORMANCE OF XGBOOST.

Evaluation set	Accuracy	AUC	Recall (poor)
Test set (20%)	87.3%	94.9%	90.7%
Entire dataset	96.1%	98.7%	98.1%

B. Feature Selection

Using RFE, a total of 35 features are selected and retained for subsequent model training. Because the final model focuses on a curated set of clinically meaningful features rather than the full laboratory profile, it remains feasible for use in distributed or bandwidth-limited network settings, where transferring complete donor datasets between HSCT centers may be inefficient.

C. Model

For classification, we employ the Extreme Gradient Boosting (XGBoost) algorithm, an ensemble method based on gradient boosting known for its high predictive accuracy, robustness, and computational efficiency [5].

III. RESULTS

A. XGBoost Performance

As shown in Fig. 2, XGBoost achieves high classification accuracy. On the 20% testing set (252 donors), 103 of 123 poor mobilizers and 117 of 129 good mobilizers are correctly classified, with only 32 misclassifications in total. On the entire dataset (799 donors), 152 of 171 poor and 616 of 628 good mobilizers are correctly classified, confirming the model's robustness.

Table I summarizes the overall performance. On the testing set, the model achieves an accuracy of 87.3%, an AUC of 94.9%, and a recall of 90.7% for poor mobilizers. On the entire dataset, the accuracy, AUC, and recall further improve to 96.1%, 98.7%, and 98.1%, respectively. These results demonstrate reliable discrimination, with particularly high recall for poor mobilizers—clinically important since missing poor donors leads to failed mobilization and unnecessary costs.

B. Explainability Analysis

1) Global explanation with SHAP: As shown in Fig. 3, SHAP summary plot ranks features by average impact, with each point representing a donor: positive SHAP values indicate

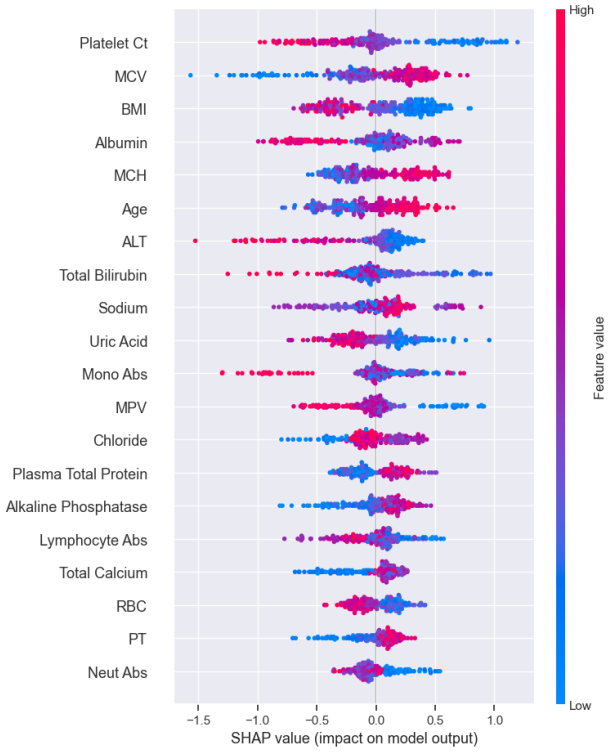

Fig. 3. SHAP summary plot.

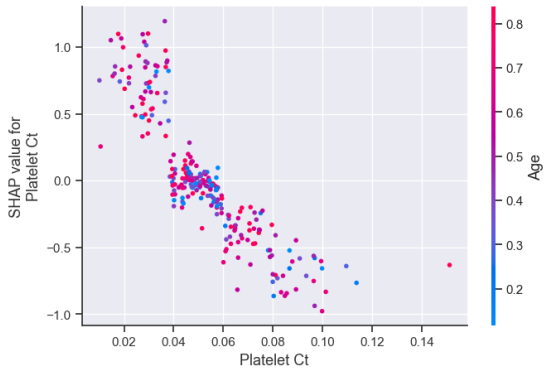

Fig. 4. SHAP dependence plot for Platelet count, colored by Age.

higher risk of poor mobilization, while negative values indicate good mobilization.

The five most influential predictors are Platelet count, MCV, BMI, Albumin, and MCH. For example, older age is associated with poor mobilization, in line with clinical knowledge that hematopoietic stem cell yield declines with donor age. Higher Platelet counts are linked to successful mobilization, consistent with established clinical evidence that platelet levels strongly predict CD34+ yield [6]. In contrast, MCV shows a novel association with poor mobilization, highlighting its exploratory value for future study. These findings are consistent with clinical knowledge, suggesting that the model captures biologically plausible risk factors in stem cell mobilization.

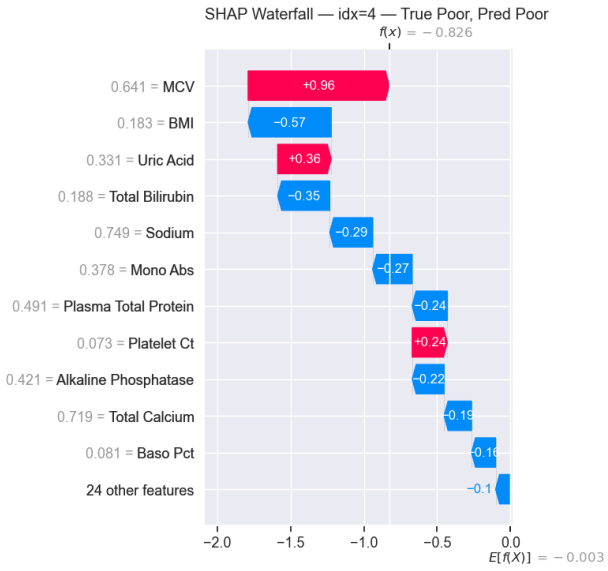

Fig. 5. SHAP waterfall plot illustrating feature contributions.

2) Feature-level explanation (SHAP dependence plots): Fig. 4 shows the SHAP dependence plot for Platelet count, with color indicating Age. Lower platelet counts are associated with positive SHAP values, shifting predictions toward poor mobilizers, while higher platelet counts decrease the probability of poor mobilization. The color gradient further suggests that older donors (red points) remain at higher risk even with moderate platelet levels, whereas younger donors (blue points) are more likely to mobilize well.

3) Local explanation for an individual donor: As shown in the SHAP summary plot (Fig. 3), higher Platelet count and Uric acid generally reduce the likelihood of poor mobilization. However, the SHAP waterfall plot for the donor in Fig. 5 illustrates how individual risk profiles can deviate from global trends. For this donor, relatively low Platelet count, low Uric acid, and high MCV are the strongest positive contributors, increasing the probability of poor mobilization. In contrast, BMI and Total Bilirubin show negative contributions, lowering this probability. This comparison highlights how SHAP links population-level patterns with case-specific explanations, providing both biological plausibility and clinically interpretable insights.

IV. Discussion

The results indicate that explainable ML can support early identification of poor mobilizers in the HSCT workflow. SHAP offers transparent and clinically interpretable insights, addressing a major limitation of previous black-box prediction approaches. In addition, relying on selected clinically meaningful features suggests that the model can be deployed efficiently within network-based or multi-center HSCT environments, where rapid sharing of interpretable predictions is valuable. Although the model performs well, future work may include

external validation or subgroup-level robustness analysis to further strengthen clinical reliability.

V. Conclusion

This paper has proposed an explainable ML framework for predicting poor stem cell mobilizers in HSCT donors. The XGBoost model achieved high accuracy with strong recall for poor mobilizers, minimizing missed high-risk cases. SHAP confirmed known predictors such as Platelet count and Age, while highlighting exploratory factors like MCV and Albumin. These findings demonstrated both biological plausibility and novel insights, supporting the clinical adoption of interpretable ML in donor screening. Given its interpretability and reliance on a curated set of clinically meaningful features rather than the full laboratory profile, the proposed approach is also suitable for integration into distributed HSCT networks, telemedicine systems, and cloud-edge clinical infrastructures, supporting reliable donor assessment across connected healthcare environments.

References

[1] C. Hosing, R. M. Saliba, S. Ahlawat, et al., "Poor hematopoietic stem cell mobilizers: a review," Ther. Adv. Hematol., vol. 5, no. 3, pp. 97–106, 2014.

[2] C. L. Morris, E. Siegel, B. Barlogie, et al., "Mobilization of CD34+ cells in elderly patients with multiple myeloma: influence of age, prior therapy, platelet count and mobilization regimen," Br. J. Haematol., vol. 120, no. 3, pp. 413–423, Feb. 2003.

[3] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," Nat. Med., vol. 25, pp. 44–56, Jan. 2019.

[4] J. Xiang, M. Shi, M. A. Fiala, F. Gao, M. P. Rettig, G. L. Uy, M. A. Schroeder, K. N. Weilbaecher, K. E. Stockerl-Goldstein, S. Mollah, and J. F. DiPersio, "Machine learning–based scoring models to predict hematopoietic stem cell mobilization in allogeneic donors," Blood Adv., vol. 6, no. 13, pp. 3822–3832, Jul. 2022.

[5] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD), pp. 785–794, 2016.

[6] Chen Y, Gao S, Wang Y, Lu M, Chu B, Shi L, Xiang Q, Fang L, Ding Y, Wang M, Liu X, Zhao X, Sun K, Bao L. "Pre-mobilization platelet count predicts stem cell yield during mobilization in patients with multiple myeloma," Cancer Pathog Ther., vol. 1, no. 1, pp. 40-45, Dec. 2022.