MACHINE LEARNING-BASED PREDICTION OF RH ALLOIMMUNIZATION RISK IN MULTIPLY TRANSFUSED B-THALASSEMIA PATIENTS: A PREDICTIVE MODELING APPROACH

Senesh Fitzroy[1]*, Thyagi Luvishewa[2], Rasika Ranaweera[1], Udeshika Yapa Bandara[2]

[1]Department of Software Engineering & Computer Security, Faculty of Computing, NSBM Green University, Homagama, Sri Lanka

[2]Department of Life Science, Faculty of Science, NSBM Green University, Homagama, Sri Lanka

*Corresponding author: dsfmendis@students.nsbm.ac.lk

Abstract

Rh antibodies predominate among alloantibodies in multiply transfused β-thalassemia patients and persist despite Rh-matched transfusion, plausibly through serologically undetected RHD/RHCE variants, yet sensitization is still detected only reactively. Conventional alloimmunization prediction designs classify patients as ever versus never alloimmunized, confounding follow-up duration with risk. This study develops an interpretable machine-learning framework that estimates each patient's 12-month Rh alloimmunization risk and updates it at every routine antibody screen. In a dynamic landmark design, predictors are recomputed at each Rh-negative screen using only data recorded on or before that date, and the outcome is a first anti-D, anti-C, anti-c, anti-E or anti-e alloantibody within 12 months. Ten prespecified candidate parameters comprise cumulative and recent transfusion burden, age at first transfusion, time-dependent splenectomy, sex, C-reactive protein, ferritin, prior non-Rh alloantibody, number of absent Rh antigens and, where available, RH variant genotype, with a serology-only variant for settings without genotyping. Riley criteria require at least 1,131 patients (approximately 129 events). Logistic regression, random forest, gradient boosting and support vector machines are compared using patient-grouped nested cross-validation, discrimination, calibration, decision-curve analysis and SHAP interpretability, following TRIPOD+AI. A tested analysis pipeline has been implemented. In a simulation study of 30 cohorts with known ground truth, the static design yielded a spurious significant association for cumulative transfusion burden in all cohorts despite a true null effect and, applied to 12-month prediction, overestimated risk approximately 13-fold (14.8% predicted versus 1.2% observed), whereas the dynamic design recovered the null effect and was well calibrated (observed-to-expected ratio 0.99). Real-patient data collection is pending. To our knowledge, this is the first Rh-restricted, thalassemia-specific alloimmunization prediction framework, and it provides a bias-resistant foundation for proactive, resource-sensitive risk stratification.

Keywords: Rh alloimmunization, β-thalassemia, machine learning, dynamic risk prediction, transfusion medicine