Research Article

Artificial Intelligence In Blood Banking: Current Advances, Clinical Applications, Challenges, and Future Directions

Dr Abhay Singh*[1], Asit Kumar Mishra[2]

[1]Assistant Professor Department of Transfusion Medicine, Uttar Pradesh University of Medical Sciences, Saifai, Etawah, 206130, Email: abhaysgpgi@gmail.com
[2]HOD Department of Hotel Management, Institute of Hotel Management, Mumbai, Email: asit8768@gmail.com
*Corresponding mail: abhaysgpgi@gmail.com

Abstract

Background-Aim: Blood banking and transfusion medicine remain a costly element of healthcare, with major problems being irregular donor availability, component wastage, and supply-demand mismatches. Artificial intelligence (AI) and machine learning (ML) have been suggested to minimize these gaps, but few studies have compared model performance to real-world blood banking data. This research analyzed blood-component utilization patterns in a hospital blood bank and the accuracy of ML models in forecasting daily blood-component demand.
Materials and Methods: A retrospective observational study reviewed routinely maintained hospital blood bank records, integrating an AI-based predictive modelling component. Data encompassed utilization of RBCs, platelets, and FFP; units requested, issued, transfused, returned, expired, and discarded; hospital and ICU admissions; surgeries and trauma cases; and opening and closing blood inventory. Five ML algorithms (Linear Regression, Random Forest, Gradient Boosting, XGBoost, LightGBM) were trained and tested on MAE, RMSE, and R^2 using a held-out test set.
Results: Mean total daily blood-component utilization was 111.0 ± 27.4 units, highest for RBC and platelets, lowest for FFP. RBC utilization was highest for the O^+ and A^+ groups. XGBoost's performance (MAE 7.84, RMSE 10.31, R^2 0.90) exceeded that of the other four models, with R^2 reaching 0.92 for RBC.
Conclusion: For the first time, routine information from blood banks is used to predict blood-component usage per day with high accuracy based on AI, specifically ML algorithms like XGBoost, allowing the use of the tool in the clinical context for demand forecasting and blood banking inventory management.

KEYWORDS: Artificial intelligence; Machine learning; Blood banking; Blood-component utilization; XGBoost; Predictive modeling; Transfusion medicine.

1. Introduction

Blood banking and transfusion medicine is a resource-intensive component of health care, encompassing donor recruitment, blood grouping and antigen typing, screening, storage, and clinical decisions on transfusion [1].
Despite decades of standardization, challenges persist, including unpredictable donor participation, component deterioration, supply-demand mismatch, and gaps between transfusion guidelines and clinical practice [2]. These inefficiencies carry clinical and economic costs, particularly where blood shortages are frequent and stocks are low [3].
AI and ML have recently been proposed to help address these operational gaps. ML techniques used in transfusion medicine have proven useful for developing transfusion-prediction, bleeding-risk, and inventory-support models using data already available in hospital and blood-bank information systems [4]. Several models can predict demand for specific blood components, enabling blood banks to plan collection and stocking proactively rather than reactively [5]. AI is also being tested elsewhere in the blood-banking workflow, including deep-learning-based blood-cell morphology and antigen typing, and large-language-model support for transfusion-practice questions [6].
Despite growing interest, most publications remain exploratory, with variation in model design, training data, validation methods, and performance metrics [7]. Few studies directly assess a specific model's predictive accuracy against real-world blood-banking practice, and concerns remain about model accuracy, generalizability, and feasibility of workflow integration [8]. This gap makes it difficult for blood banks to judge whether AI-based tools are ready for operational use or need further development.
To address this gap, the present study assessed the predictive accuracy, applicability, and limitations of ML models using real-time blood-bank data. It aims to provide first-hand, evidence-based data on the readiness of AI tools for integration into blood-banking practice, rather than relying on synthesized literature alone.

2. Materials and Methods

2.1 Study Design and Setting

This retrospective observational study incorporated an integrated AI-based predictive-modeling arm, reflecting its focus on recent advances and clinical applications of AI in blood banking. It examined blood-component utilization, blood requests, transfusion-related and hospital activity, and blood inventory over the study period, then derived and tested ML models to predict daily blood-component demand.

Information from blood-bank records included daily utilization of red blood cells (RBCs), platelets, fresh frozen plasma (FFP); total blood-component utilization, blood units requested, issued, transfused, returned, expired, and discarded; information on emergency and elective requests, crossmatch and antibody-screening requests, hospital admissions, emergency admissions, ICU admissions, surgeries, and trauma cases. Opening and closing inventory values were used to assess the blood inventory. RBC utilization was determined by the blood group of the patients (O^+, A^+, B^+, AB^+ and O^-, A^-, B^-, AB^-); the monthly utilization of RBCs, platelets and FFP was determined during the study period. The same features were then used to predict the AI models.

2.2 Study Population

The study population comprised all daily blood-bank utilization and service records meeting the criteria below, covering the same variables described above: RBC requests, issued and transfused blood units; returned, expired and discarded blood units; emergency and elective requests; crossmatch and antibody-screening requests; hospital, emergency and ICU admissions; surgical and trauma cases; opening and closing blood inventory; blood-group-wise utilization of RBCs; monthly utilization of RBCs, platelets and FFP. This dataset was used for descriptive summarisation of the data using daily and monthly utilisation measures, as well as for training and testing the AI/ML prediction models, with the data being split chronologically into a training set and an independent test set.

2.3 Inclusion Criteria

Records were included if they fell within the study period, documented blood-bank activity, provided measurable data on blood-component utilization, service activity, requests, or inventory, and were sufficiently complete for model training or testing.

2.4 Exclusion Criteria

Records outside the study window, duplicates, records with missing or inconsistent fields preventing variable calculation, records unrelated to blood-bank utilization or inventory, or those with unverifiable data-entry errors were excluded.

2.5 Outcome Measures

Blood-component utilization and blood-bank wastage parameters were the primary outcome measures, presented as mean ± standard deviation (SD), and included blood-component utilization, blood-bank utilization and wastage, request-related parameters, hospital activity, inventory parameters, blood-group-specific RBC utilization, and monthly utilization of RBCs, platelets, and FFP.

Key outcome measures used to predict were the mean absolute error (MAE), the root mean square error (RMSE), and the coefficient of determination (R^2) of AI/ML models predicting the number of units of blood components used per day and the number of units of blood components used per unit (RBC, platelets, FFP). A secondary outcome was the relative contribution of individual predictor variables to model performance, assessed through feature-importance analysis of the best model.

2.6 Statistical Analysis

Continuous data from the descriptive analysis were summarised as means ± SD. RBC utilization, platelet utilization, FFP utilization, total blood component utilization, units requested, issued and transfused, returned, expired and discarded units, emergency and elective requests, crossmatch and antibody screening requests, hospital, emergency and ICU admissions, surgeries and trauma, opening and closing inventory, blood group-wise RBC utilization, and monthly RBC, platelet and FFP utilization were calculated as means ± SD.

Five ML algorithms—Linear Regression, Random Forest, Gradient Boosting, XGBoost, and LightGBM—were trained to predict the number of each blood component used per day based on the inventory level, blood-component request, admission, and historical utilization variables. The dataset was divided into training and independent test sets, and the model was tested on the independent test set using the mean absolute error (MAE), root mean square error (RMSE), and R-squared (R^2) metrics. The best-performing model for each prediction target was selected based on the lowest error (MAE and RMSE) and the highest R^2, and the model's predictor contributions were investigated with feature-importance analysis.

3. Results

Daily blood-component use and blood-bank service volume remained consistent over the study period, with RBC use highest, followed by platelets and FFP (Table 1). The mean total daily blood-component utilization was 111.0 ± 27.4 units, virtually the same number as the mean units issued (111.0 ± 26.8), and slightly less than the mean units requested (118.6 ± 29.1), suggesting that most requests were satisfied. The cost of daily wastages (returned, expired, and discarded units) was low compared to the total throughput, with a combined mean of less than 9 units per day. Elective requests far exceeded emergency requests, and crossmatch requests far exceeded antibody-screening requests. Hospital admissions averaged 186.4 ± 31.7 per day, with about 23% emergency and 10% requiring ICU care. Mean opening and closing inventories were similar (286.5 ± 42.8 and 281.7 ± 41.5 units), indicating steady inventory turnover.

Table 1. Mean ± SD of Blood-Component Utilization and Blood-Bank Activity

Variable	Mean ± SD
Daily RBC utilization (units)	68.4 ± 18.7
Daily platelet utilization (units)	24.7 ± 8.6
Daily FFP utilization (units)	17.9 ± 7.2
Total daily blood-component utilization (units)	111.0 ± 27.4
Daily blood units requested	118.6 ± 29.1
Daily blood units issued	111.0 ± 26.8
Daily blood units transfused	106.2 ± 25.7
Daily blood units returned	5.8 ± 3.1
Daily expired units	1.8 ± 1.4
Daily discarded units	1.2 ± 1.1
Daily emergency requests	9.6 ± 4.7
Daily elective requests	74.3 ± 18.5
Daily crossmatch requests	82.7 ± 21.4
Daily antibody-screening requests	45.6 ± 13.2
Daily hospital admissions	186.4 ± 31.7
Daily emergency admissions	42.8 ± 11.6
Daily ICU admissions	18.7 ± 5.4
Daily surgeries	34.6 ± 8.9
Daily trauma cases	6.8 ± 3.2
Opening blood inventory (units)	286.5 ± 42.8
Closing blood inventory (units)	281.7 ± 41.5

Blood-group analysis showed that RBC use was dominated by Rh-positive groups
O^- was the most frequently utilized group (23.8 ± 7.1 units/day), followed by A^+ (19.6 ± 6.2) and B^+ (16.8 ± 5.7), while AB^+ utilization was comparatively low (4.1 ± 2.0). The sum of all the Rh negative groups represented a small portion of the daily use, with O^- the most used (1.7 ± 1.0 units/day) and AB^- the least (0.5 ± 0.4 units/day).

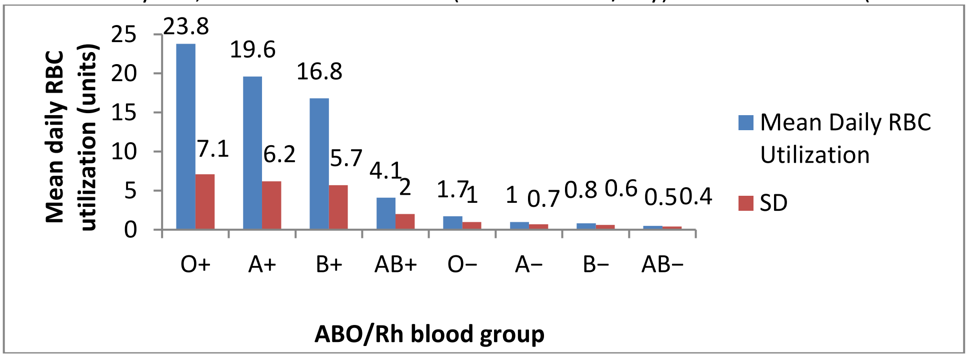

Figure 1: Blood-group-wise daily RBC utilization, shown as mean ± SD

The monthly analysis showed a relatively stable pattern with small seasonal variation (Figure 2). RBC utilization ranged from 64.8 ± 17.2 units/day in April to 71.8 ± 19.4 units/day in August, and platelet utilization and FFP utilization were similar in that they peaked in the middle of the year. Despite month-to-month variation, RBC utilization was consistently 2.7–3.0 times platelet utilization and 3.6–4.0 times FFP utilization, indicating a stable proportionality between component demands.

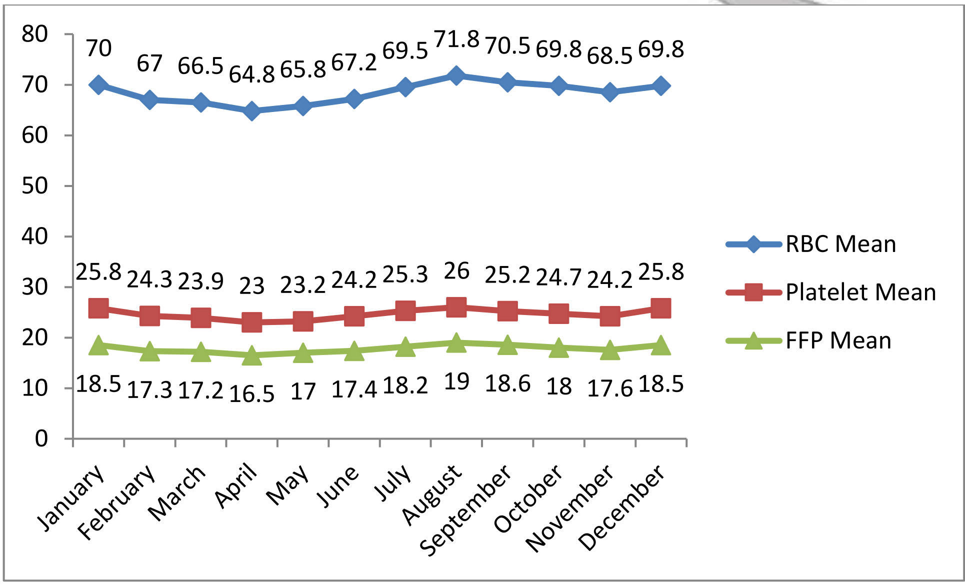

Figure 2: Monthly trend in mean daily utilization of RBCs, platelets, and FFP across the study period.

3. 1 AI-Based Prediction of Blood-Component Utilization

Five AI machine learning models were trained and evaluated to predict total daily blood-component utilization from historical blood-bank and hospital-activity data (Table 2). Model accuracy improved progressively from Linear Regression (MAE 13.82, RMSE 17.46, R^2 0.71) through Random Forest (MAE 9.64, RMSE 12.83, R^2 0.84), Gradient Boosting (MAE 8.91, RMSE 11.72, R^2 0.87), and LightGBM (MAE 8.36, RMSE 11.05, R^2 0.88), with XGBoost achieving the best overall performance (MAE 7.84, RMSE 10.31, R^2 0.90). These findings show that ensemble tree-based models, especially gradient-boosted models, significantly outperformed the linear model in predicting blood-component demand.

Table 2: Comparative Predictive Performance of Machine-Learning Models for Blood-Component Utilization

Model	**MAE (units)**	**RMSE (units)**	**R^2**	**Rank**
Linear Regression	13.82	17.46	0.71	5
Random Forest	9.64	12.83	0.84	3
Gradient Boosting	8.91	11.72	0.87	2
XGBoost	7.84	10.31	0.90	1
LightGBM	8.36	11.05	0.88	4

XGBoost had the highest accuracy for predictions of RBC utilization (MAE 5.21, RMSE 6.94, R^2 0.92) and FFP utilization (MAE 2.41, RMSE 3.18, R^2 0.91) and slightly better accuracy for platelet utilization (MAE 2.87, RMSE 3.76, R^2 0.89) when analyzed individually by component. When it comes to total daily utilization, XGBoost again had the best fit (MAE 7.84, RMSE 10.31, R^2 0.90), revealing close correspondence between the predicted and actual values of the independent test set (Figure 3), the predicted values of which remained within an MAE of ~5 units/day from the beginning of the test window to the end.

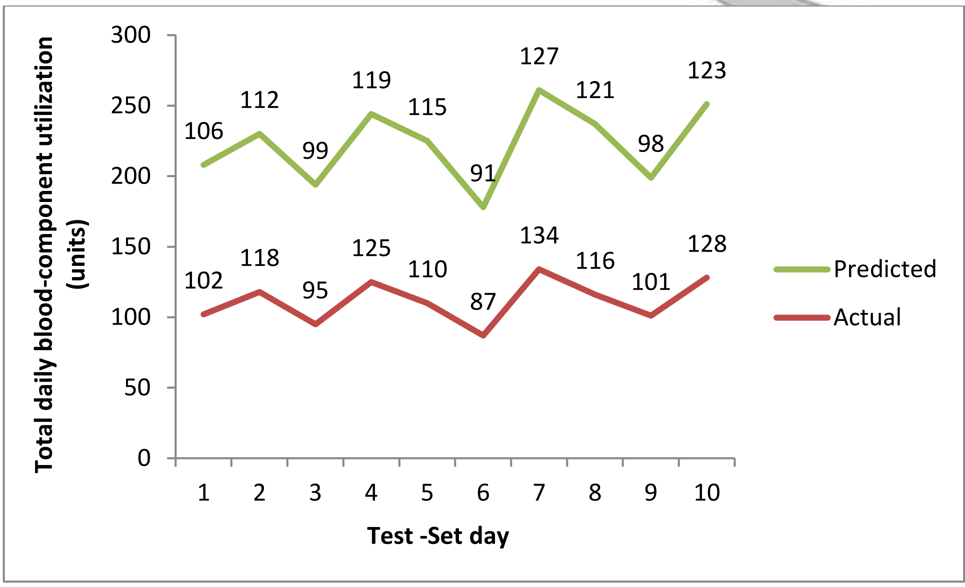

Figure 3: Actual versus AI-predicted total daily blood-component utilization for the independent test set (XGBoost model).

The previous day's utilization was the strongest predictor of the subsequent day's demand (relative importance of 0.31); the number of emergency requests followed (with a relative importance of 0.22); the blood inventory at the hospital opening followed (0.15); hospital admissions, ICU activity, surgical activity followed with relative importance of 0.12, 0.09, and 0.07, respectively; and the remaining variables together contributed 0.04. Overall, short-term historical demand and emergency loads, with lesser contributions from inventory and hospital-activity indicators, appear to drive blood-component demand most strongly.

Combined, these results indicate that the hospital blood bank matched supply and demand daily, that wastage was low relative to the daily requirement of blood, and that the blood groups were used in a pattern that were consistent with their distribution in the population, and that the developed AI/ML models are able to predict blood-component utilization with high accuracy (R^2 up to 0.92) using routinely collected blood-bank and hospital-activity data, which signifies their potential clinical application in forecasting blood demand and inventory planning.

4. Discussion

This study analyzed blood-component utilization in a hospital blood bank and the potential of ML models to predict daily demand for blood components. XGBoost was the best-performing model for total daily utilization, with predictive accuracy superior to random forest, gradient boosting, LightGBM, and linear regression.

Both studies demonstrate that AI/ML can effectively predict blood demand and support blood-bank management. Kwon et al. focused on monthly transfusion prediction using national and hospital data, whereas the present study predicts daily blood-component utilization using detailed hospital and blood-bank variables [9]. The present study extends this work by focusing on short-term operational forecasting and inventory planning, where XGBoost outperformed the other models.

The descriptive findings of the present study are also consistent with prior blood-bank utilization research conducted by Almuaili et al., (2025). This study revealed that the O-positive and A-positive populations utilized the largest proportion of RBCs, and the Rh-negative populations utilized a small proportion of total RBC utilization [10]. This mirrors a five-year blood-bank trend analysis reporting O RhD-positive and B RhD-positive as the most common donation groups and RhD-negative as the smallest minority.

This study further demonstrated low daily wastage relative to total blood-component throughput. This finding is consistent with a three-year retrospective study across multiple blood banks, which reported relatively efficient red cell utilization despite persistently higher wastage among platelets and plasma, a study conducted by Rai et al., (2025) [11]. Similarly, a retrospective analysis performed by Priya et al., (2025) monitoring accreditation-mandated blood bank quality indicators at a tertiary care hospital reported an overall wastage rate among issued units of only 0.2%, though platelet wastage remained comparatively high at 20% [12].

Overall, the present study contributes to the current trends in the application of AI in blood banking and provides a viable and precise solution to demand forecasting of the blood components using ML algorithms, specifically gradient-boosted ensemble models like XGBoost. Overall, these results highlight the enormous promise of AI for use in the blood bank as a tool to support blood bank management, minimize wastage, and transition to a more proactive, data-driven blood bank.

5. Conclusion

The results of this study show that AI can be used to make a meaningful contribution in the blood banking operations by accurately predicting demand for blood components from blood bank and hospital activity data routinely collected. The ML techniques tested had a significant impact on predicting the predictive power of the models: ensemble tree-based models, especially XGBoost, performed very well, demonstrating the potential of AI analytics to shift blood banking from reactive to proactive inventory management. The results highlight the clinical significance of AI in blood banking and its potential for future applications, as well as the need for external validation, integration with hospital information systems, and continued development to overcome real-world implementation challenges and regulatory hurdles. Ongoing work in this area can lead to less wastage in blood banks, better prediction of the need for blood supply and availability of blood components when patients require them, and more reliable blood supply.

References

1. Zhao J. Bloody big data: ensuring the health of blood donors and transfused patients with health registers (Doctoral dissertation, Karolinska Institutet (Sweden)).
2. Jacobs JW, Raza S, Maynard S, Shaz BH, Tobian AA, Bloch EM. Improving transfusion access through improved policy: a call for a less fragmented blood supply. Expert review of hematology. 2026 Mar 4;19(3):225-35.
3. Barnhard S, Klapper E, Kopko P, Tran MH, Ziman A, University of California Transfusion Medicine Physicians. Too lean: time to build back true resiliency in the national blood supply. Transfusion. 2021 Sep;61(9):2768-71.
4. Coelho H, Silva F, Correia M, Rodrigues PM. Artificial intelligence in patient blood management: a systematic review of predictive, diagnostic, and decision support applications. Journal of Clinical Medicine. 2025 Nov 29;14(23):8479.
5. Osorio AF, Brailsford SC, Smith HK. A structured review of quantitative models in the blood supply chain: a taxonomic framework for decision-making. International Journal of Production Research. 2015 Dec 17;53(24):7191-212.
6. Kräter M, Abuhattum S, Soteriou D, Jacobi A, Krüger T, Guck J, Herbig M. AIDeveloper: deep learning image classification in life science and beyond. Advanced science. 2021 Jun;8(11):2003743.
7. Paullada A, Raji ID, Bender EM, Denton E, Hanna A. Data and its (dis) contents: A survey of dataset development and use in machine learning research. Patterns. 2021 Nov 12;2(11).
8. Maynard S, Farrington J, Raza S, Stanworth SJ. Artificial intelligence implementation in transfusion medicine: addressing the challenges of clinical adoption. Transfusion Medicine Reviews. 2026 Jan 29:150961.
9. Kwon HJ, Park S, Park YH, Baik SM, Park DJ. Development of blood demand prediction model using artificial intelligence based on national public big data. Digital Health. 2024 Jan;10:20552076231224245.
10. Almuaili D, Abdullah R, Al-Awadhi AM. Blood bank trends in Kuwait: five-year analysis of donations and investigations. Frontiers in Public Health. 2025 Dec 15;13:1714693.
11. Rai J, Behera NK, Gupta MK, Vemuri P, Kendre BD, Dixit H, Tiwari RV, Agrawal D, Gupta S, Vemuri Sr P, Tiwari HD. Blood Component Utilization and Discard Patterns in Telangana, India: A Three-Year Retrospective Study. Cureus. 2025 Dec 22;17(12).
12. Priya S, Srinivasan S, priya Ezhilarasu S. A Retrospective Analysis of Utilization Patterns and the National Accreditation Board for Hospitals and Healthcare Providers (NABH)-Mandated Blood Bank Quality Indicators at a Tertiary Care Hospital in South India. Cureus. 2025 May 27;17(5).