

AI and Blood Banking: Predicting Transfusion Demand – A Systematic Review of Forecasting Approaches

Merlin Engelke^a Osman Alperen Çınar-Koraş^a Jens Kleesiek^{a, b, c}
Peter A. Horn^d

^aInstitute for Artificial Intelligence in Medicine, University Medicine Essen, Essen, Germany; ^bFaculty of Computer Science and Medical Faculty, University of Duisburg-Essen, Essen, Germany; ^cDepartment of Physics, TU Dortmund University, Dortmund, Germany; ^dInstitute for Transfusion Medicine, University Medicine Essen, Essen, Germany

Keywords

Machine learning · Blood transfusion · Demand prediction · Forecasting · Platelets

Abstract

Introduction: Accurate prediction of blood product demand is essential for maintaining adequate supply while minimizing wastage. Artificial intelligence (AI) and machine learning (ML) approaches have emerged as promising tools for transfusion demand forecasting across multiple scales, ranging from individual patients to regional supply chains. This review provides a structured overview of AI applications in blood banking, organized by prediction scope and clinical objective. **Methods:** We conducted a systematic review following PRISMA 2020 guidelines, searching PubMed and Scopus for studies published between January 2015 and December 2025. Studies applying AI or ML to transfusion demand prediction were included. We organized findings into a four-tier framework based on prediction scope and clinical goal and systematically compared technical approaches. **Results:** Out of 535 identified records, supplemented by a targeted search during peer review, 24 studies met inclusion criteria. The four-tier framework revealed distinct prediction challenges: (1) patient-individual prediction (goal: improve individual care quality) achieved areas under the curve (AUCs) of 0.45–0.97 across five studies for 24-h to 72-h prediction, (2) procedure-specific models (goal: optimize surgical blood ordering) achieved AUCs of 0.70–0.91 across six studies, (3) facility-level optimization (goal: reduce

wastage and shortages) approximately halved platelet outdating rates and achieved mean absolute percentage errors as low as 4.18% across eight studies for daily to weekly forecasting, and (4) regional forecasting (goal: secure long-term supply) achieved R^2 values up to 0.85 across five studies for monthly forecasting. Deep learning excelled for temporal patterns in individual patients, while traditional statistical methods (ARIMA, SARIMA) proved effective for aggregated regional forecasting, where deep learning approaches showed no consistent advantage over simpler models. Prediction horizons varied by blood product: 3–7 days for platelets versus weekly to monthly for red blood cells. **Conclusion:** AI-driven demand prediction offers substantial potential to improve blood supply management across all tiers of the blood supply chain. The choice of modeling approach should be guided by the prediction scope, available data infrastructure, and specific clinical objectives. Standardized external validation remains critical for clinical deployment.

© 2026 The Author(s).

Published by S. Karger AG, Basel

Introduction

Accurately estimating how much blood an individual patient will need before a surgical or therapeutic intervention is central to safe transfusion practice. Overestimation leads to unnecessary cross-matching and wastage of scarce products, while underestimation risks delayed

availability and adverse patient outcomes [1]. This challenge extends beyond the individual bedside: at institutional and regional levels, the management of blood product supply requires balancing adequate availability against wastage of perishable resources. Platelets have a shelf life of only 5–7 days, while packed red blood cells (pRBCs) expire after 35–42 days [2]. This limited viability, combined with the stochastic nature of both supply (donor availability) and demand (patient needs), creates complex inventory management problems.

The consequences of suboptimal blood supply management are significant. Shortages can delay surgeries, compromise patient outcomes, and in extreme cases lead to preventable deaths [3, 4]. Conversely, overstocking leads to wastage (outdating rates for platelets range from 5 to 20% in many institutions) [5], and prolonged storage has been linked to adverse transfusion outcomes [6]. In an era of declining donor populations and aging demographics that increase transfusion demand, optimizing this balance has become increasingly critical [7, 8].

Artificial intelligence (AI) and machine learning (ML) have emerged as powerful tools for addressing this challenge. By learning patterns from historical data, these approaches can forecast transfusion needs with greater accuracy than traditional methods [9]. Applications range from predicting which individual patients will require transfusion within the next 24–72 h to forecasting regional blood demand months in advance.

This systematic review synthesizes the current evidence on AI and advanced statistical approaches for blood transfusion demand prediction. We organized findings into four tiers based on prediction scope and clinical objective:

1. Patient-individual prediction: improve individual patient care quality through personalized transfusion management
2. Procedure-related prediction: optimize care structures such as Maximum Surgical Blood Order Schedules (MSBOS) and resource allocation for specific interventions
3. Facility/inventory management: optimize institutional blood bank operations, reducing both wastage and shortages
4. Regional/national forecasting: secure long-term blood supply at population level through strategic planning

While presented as distinct tiers, these prediction levels are inherently interconnected. Patient-individual predictions (tier 1) aggregate into facility-level demand (tier 3), which in turn feeds regional supply planning (tier 4). Procedure-specific models (tier 2) directly inform institutional ordering policies that shape inventory requirements. We systematically compare the technical approaches employed at each tier, discussing their respective advantages and limitations, appropriate prediction horizons, and the critical distinction between platelet and red blood cell forecasting.

Methods

Search Strategy and Information Sources

This systematic review was conducted following PRISMA 2020 guidelines [10]. We searched PubMed and Scopus from January 2015 to December 2025. The search was restricted to title and abstract fields using the following search string:

(“artificial intelligence” OR “machine learning” OR “deep learning” OR “neural network*” OR “random forest” OR “gradient boosting” OR “XGBoost” OR “LSTM” OR “ARIMA”) AND (“transfusion” OR “blood bank*” OR “blood product*” OR “blood supply” OR “blood component*”) AND (“prediction model” OR “predictive model” OR “predict transfusion” OR “predicting transfusion” OR “transfusion prediction” OR “transfusion risk” OR “transfusion requirement*” OR “transfusion need*” OR “forecast*” OR “transfusion demand” OR “blood demand” OR “RBC demand” OR “red blood cell demand” OR “transfusion receipt” OR “receiving transfusion” OR “inventory” OR “supply chain”)

Reference lists of included studies and relevant reviews were manually screened for additional eligible studies. During peer review, a supplementary search was conducted to address potential gaps in the original search string. Additional terms, specifically “platelet,” “wastage,” “shortage,” and “outdate,” were searched in combination with the original AI/ML methodology terms. This supplementary search identified two additional eligible studies [11, 12] that were not captured by the original query, bringing the total to 24 included studies.

Selection Process

Titles and abstracts were screened independently by M.E. and O.A.C.-K. Full-text articles were assessed for eligibility by M.E. and O.A.C.-K., with disagreements resolved through discussion with J.K. and P.A.H.

Data Collection Process

Data extraction was performed by M.E. and O.A.C.-K. using a standardized form. The extraction process is illustrated in Figure 1. Extracted variables included study design, sample size, prediction tier, technical approach, performance metrics (area under the curve [AUC], area under the receiver operating characteristic curve [AUROC], mean absolute percentage error [MAPE], root mean-squared error [RMSE], R^2), validation strategy, and key predictors.

Eligibility Criteria

Studies were included if they (1) applied AI/ML or advanced statistical methods (ARIMA, exponential smoothing) to transfusion demand prediction or forecasting at patient, procedure, facility, or regional level; (2) reported quantitative performance metrics (AUC, AUROC, MAPE, RMSE, R^2 , sensitivity/specificity); (3) were original research articles; and (4) were published in

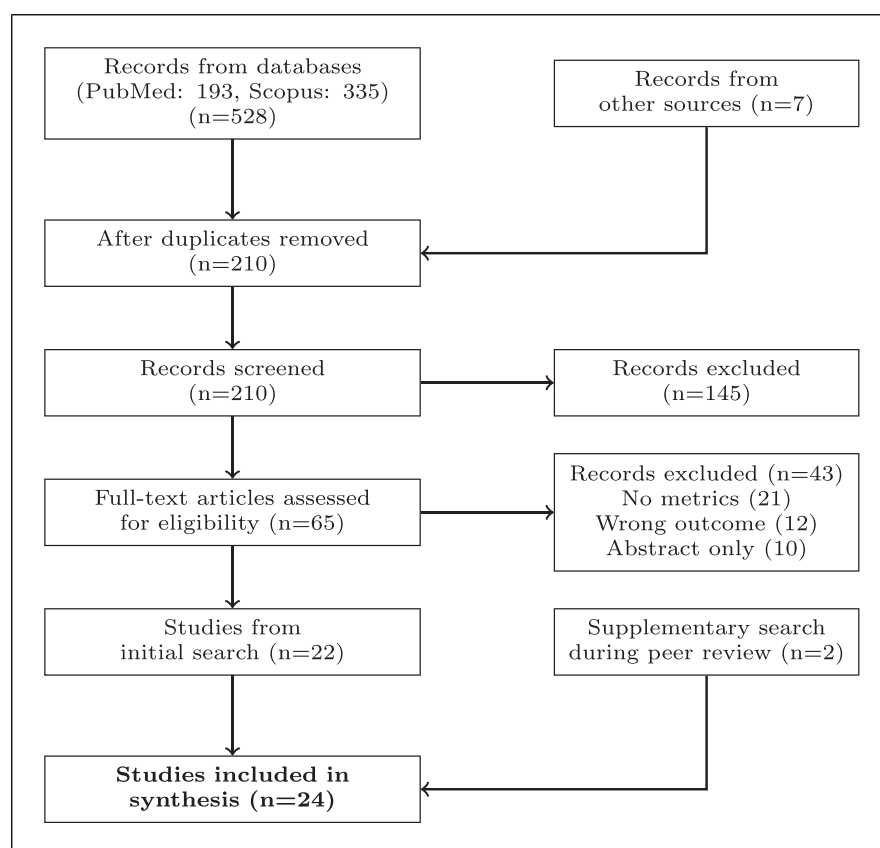


Fig. 1. PRISMA 2020 flow diagram of study selection. Of 535 initially identified records (528 from databases, 7 from other sources), 210 remained after duplicate removal. After title/abstract screening (145 excluded) and full-text assessment (43 excluded: 21 no metrics, 12 wrong outcome, 10 abstract only), 24 studies were included for synthesis.

English with full text available. Exclusion criteria comprised conference abstracts without full papers; studies without quantitative metrics; non-predictive AI applications (blood typing, antigen matching, donor screening); massive transfusion protocol activation prediction without demand quantification; bleeding risk prediction without transfusion outcomes; and reviews, editorials, or commentaries.

Risk-of-Bias Assessment

Risk of bias was assessed using the Prediction model Risk of Bias Assessment Tool (PROBAST) [13], evaluating four domains: participants, predictors, outcome, and analysis. All authors assessed each study independently. Detailed PROBAST assessments for all 24 included studies and the PRISMA 2020 checklist are provided in the online supplementary Material (for all online suppl. material, see <https://doi.org/10.1159/000552969>).

Reporting Bias and Certainty Assessment

Formal assessment of reporting bias (e.g., funnel plots, Egger's test) was not performed due to the heterogeneity of outcomes and lack of comparable effect sizes across studies. Certainty of evidence assessment using GRADE [14] or similar frameworks was not conducted as the included studies were prediction model development and validation studies rather than intervention trials.

Synthesis Methods

Due to heterogeneity in study populations, outcomes, and methods, we performed narrative synthesis organized by our four-tier framework rather than meta-analysis. Technical approaches were systematically compared regarding their suitability for different prediction tasks, data requirements, and performance characteristics. Sensitivity analyses were not conducted given the narrative synthesis approach.

Results

Study Selection

Figure 1 presents the PRISMA flow diagram. Database searches identified 528 records (PubMed: 193, Scopus: 335), supplemented by 7 records from reference list screening and citation searching. After removing 325 duplicates, 210 papers were screened by title and abstract, excluding 145. Of 65 full-text articles assessed, 43 were excluded (21 no quantitative metrics, 12 wrong outcome, 10 conference abstracts only), yielding 22 studies. A supplementary search conducted during peer review with additional terms ("platelet," "wastage," "shortage," "outdate") identified 2 further eligible studies, bringing the total to 24 studies for synthesis (Table 1).

Table 1. Summary of included studies by tier

Tier	Focus area	Model	Performance	Validation	Sample size	Key study
1	Hospital admission	Gradient boosting	AUC: 0.966	Internal	206,271	Mitterecker et al. [15]
1	ICU transfusion	Meta-learner	AUROC: 0.97	Temporal	72,072	Rafiei et al. [16]
1	Lumbar fusion	XGBoost	AUC: 0.958	Internal	1,905	Lang et al. [17]
1	Patient individual	LSTM	AUC-PR: 0.84	Internal	34,809	Engelke et al. [18] (2023a)
1	Patient individual (3d)	ML classifiers	AUC-PR: 0.45	Internal	25,190	Engelke et al. [19] (2023b)
2	Preoperative RBC	LightGBM	AUC: 0.908	Internal	130,996	Feng et al. [20]
2	Cesarean section	GBDT	AUC: 0.89	Simulation	13,372	Chen H. et al. [21]
2	Knee arthroplasty	GBM	AUC: 0.880	External	2,086	Jo et al. [22]
2	Hip fracture	SVM	AUC: 0.859	Internal	1,355	Chen X. et al. [23]
2	Liver transplant	XGBoost	AUROC: 0.813	Prospective	1,193	Liu et al. [24]
2	Cardiac surgery	Logistic Reg.	ROC: 0.7	External	104,050	Flint et al. [25]
3	Platelet usage	Statistical	Expiry: 10.5%→3.2%	Internal	29 months	Guan et al. [11]
3	Blood components	SARIMAX/LSTM	R^2 : 0.876	Internal	6 months	Wang et al. [26]
3	RBC ordering	STL + XGBoost	Inventory: −40%	Temporal	227,944	Li et al. [27]
3	RBC demand	Autonomous	MAPE: 4.18–5.85%	Internal	3 years	Turkulainen et al. [28]
3	Platelet demand	ARIMA	MAPE: 18.77%	Internal	61,377	Motamedi et al. [29] (2024)
3	Platelet inventory	LASSO	MAPE: 18.11%	Simulation	45,900	Schilling et al. [30]
3	Platelet Inv. Mgmt	Data-driven	Shortage/ waste opt.	Simulation	3 years	Motamedi et al. [12] (2025)
3	Platelet supply chain	ANN	RMSE: 0.054–0.132	Internal	6 years	Fanoodi et al. [31]
4	National (Korea)	Ensemble Boost	R^2 : 0.85 [‡]	Internal	12 years	Kwon et al. [32]
4	Regional (China)	SARIMA	MAPE: 13.69%	Retrospective	10.5 years	Liu et al. [33]
4	Regional (Iran)	ARIMA	MAE-based	Internal	8 years	Sarvestani et al. [8]
4	National (Taiwan)	Seasonal ARIMA	MAPE: 80%	Internal	5 years	Shih and Rajendran [34]
4	Regional (Tehran)	MLSTM	R: 0.98/0.97	Internal	20 months	Shokouhifar et al. [35]

SVM, support vector machine. [‡]Specifically for O+ platelet concentrates.

Tier 1: Patient-Individual Prediction
Goal: Improve Individual Patient Care Quality through Personalized Transfusion Management
Five studies addressed patient-level transfusion prediction with the aim of enabling proactive clinical management. Mitterecker et al. [15] conducted the largest study in this tier, evaluating 4 ML algorithms across 206,271 hospital admissions at 3 tertiary hospitals in Western Australia; gradient boosting achieved AUC of 0.966 for predicting any RBC transfusion at admission. Rafiei et al. [16] achieved a best-scenario AUROC of 0.97 using a meta-learner ensemble combining multiple base classifiers for 24-h ICU transfusion prediction across 72,072 admissions. Lang et al. [17] employed XGBoost with SHAP (SHapley Additive exPlanations) for model interpretability in lumbar fusion procedures, attaining AUC of 0.958; however, this model

included intraoperative predictors (estimated blood loss, fluid intake volume, and operative time) that are unavailable before surgery, which severely limits its utility for preoperative transfusion prediction and means the reported performance does not reflect a clinically actionable preoperative model. Engelke et al. [18] developed Long Short-Term Memory (LSTM) networks [36] for 24-h platelet demand prediction in hematology-oncology patients, achieving an area under the precision-recall curve of 0.84. The same group reported ML classifiers achieving area under the precision-recall curve of 0.45 for 3-day individual platelet demand prediction [19].

Key findings indicate that AI performs best when disease trajectories are physiologically progressive (e.g., chemotherapy-induced thrombocytopenia) rather than stochastic (e.g., trauma). Dynamic laboratory values,

particularly platelet count trends over preceding days, consistently outperformed static demographics as predictors [15, 16]. The exceptional ICU accuracy (AUROC 0.97) reflects the value of high-frequency vital sign monitoring and laboratory data in enabling precise short-term predictions. Notably, the dominance of dynamic laboratory values at the individual level connects directly to tier 2, where preoperative hemoglobin, itself a single-timepoint laboratory value, emerges as the strongest predictor, suggesting that even simplified snapshots of the physiological signals driving Tier 1 models carry substantial predictive power for surgical decision-making.

Tier 2: Procedure-Related Prediction

Goal: Optimize Care Structures and Surgical Blood Ordering (e.g., MSBOS Guidelines)

Six studies focused on surgical transfusion prediction with implications for Maximum Surgical Blood Order Schedules (MSBOS). Feng et al. [20] conducted the largest study in this tier, applying LightGBM to 130,996 surgical patient records for preoperative RBC demand prediction, achieving AUC of 0.908 and outperforming clinician judgment for 85% of cases. Chen H. et al. [21] developed a gradient boosting model for predicting RBC transfusion in cesarean sections across 13,372 cases (10,147 for model development and 3,225 for simulation prediction), achieving AUC of 0.89. Jo et al. [22] developed gradient boosting models for total knee arthroplasty achieving AUC of 0.880, identifying preoperative hemoglobin and tranexamic acid use as primary predictors. Chen X. et al. [23] applied support vector machine classification to hip fracture surgery with a testing AUC of 0.859, with hemoglobin and age as key variables. Liu et al. [24] employed XGBoost for liver transplantation surgery, achieving AUROC of 0.813 with prospective validation on an independent cohort. Flint et al. [25] developed the Australian Cardiac Surgery Platelet Transfusion (ACSePT) risk prediction tool using multivariate logistic regression on 91,521 cardiac surgery patients and validated on an independent temporal cohort of 12,529 patients, with the full model achieving ROC of 0.70 and the simplified ACSePT scoring tool achieving ROC of 0.67.

Preoperative hemoglobin emerged as the single strongest predictor across all surgical studies, consistent with established transfusion triggers [37]. Performance varied substantially with surgical complexity: models for relatively standardized procedures (arthroplasty, cesarean section) achieved AUCs above 0.88, while Flint et al.'s [25] comparatively modest ROC of 0.70 highlights the challenge of predicting outcomes in complex cardiac procedures with multiple confounders including cardiopulmonary bypass duration, surgical complexity, and intraoperative blood loss variability. Of the six tier 2

studies, only Jo et al. [22] and Flint et al. [25] performed external validation, while Liu et al. [24] used prospective validation; the remaining three relied on internal validation only. These externally validated models provide more realistic performance estimates than internally validated studies. This validation gap is not unique to tier 2. Across all 4 tiers, only 4 of 24 studies employed external, multi-site, or prospective validation, representing the most critical limitation of the current evidence base and a pattern that recurs throughout this review.

Tier 3: Facility/Inventory Management

Goal: Optimize Institutional Blood Bank

Operations, Reducing Wastage and Shortages

Eight studies addressed facility-level optimization with direct operational implications. Guan et al. [11] developed a statistical forecasting model for platelet usage at Stanford University Medical Center, using 29 months of institutional data to identify that platelet consumption depends strongly on weekday/weekend patterns, hospital census, and the number of patients with abnormal blood counts; the model reduced platelet expiration rates from 10.5% to 3.2%, with the authors estimating potential annual savings of approximately USD 80 million if implemented across US transfusion services. Schilling et al. [30] demonstrated that LSTM and LASSO (Least Absolute Shrinkage and Selection Operator) models could reduce platelet outdating from approximately 10%–5%, translating to approximately USD 250,000 annual savings at a single institution. Li et al. [27] combined ML forecasting with operations research optimization using integer programming, achieving 40% inventory reduction and 60% decrease in order frequency while maintaining service levels. Turkulainen et al. [28] developed an autonomous forecasting system that automatically selects among 11 methods (including ARIMA, exponential smoothing, and neural networks) for weekly RBC demand prediction, achieving MAPE of 5.85% (Finland) and 4.18% (Netherlands) on real demand data. Wang et al. [26] integrated SARIMAX and LSTM models for daily blood component demand forecasting, achieving R^2 of 87.6% for apheresis platelets and MAPE as low as 0.37%. Motamedi et al. [29] compared univariate (ARIMA, Prophet) versus multivariate (LASSO, random forest, LSTM) approaches for platelet forecasting across 61,377 transfusions, finding that with sufficient historical data, the simpler ARIMA model achieved comparable performance to multivariate approaches, with a best MAPE of 18.77%. In a companion study, Motamedi et al. [12] extended this work to inventory management, demonstrating that incorporating data-driven demand forecasts into ordering policies effectively minimizes both shortages and waste, with the system tolerating forecast errors better as demand is aggregated across facilities.

Fanoodi et al. [31] compared artificial neural networks (ANN) and ARIMA models for platelet supply chain forecasting in Iran, with ANN achieving RMSE values of 0.054–0.132 across eight platelet types over 6 years of daily demand data.

These findings support a tiered implementation strategy: smaller hospitals with limited data infrastructure can effectively use ARIMA or exponential smoothing models, while tertiary centers with rich clinical data warehouses benefit from LSTM or gradient boosting approaches. The autonomous method selection approach of Turkulainen et al. [28] offers a practical middle ground, automatically adapting to local demand patterns without requiring ML expertise. The integration of demand prediction with operations research optimization, as demonstrated by Li et al. [27] and Motamedi et al. [12], represents the cutting edge of this tier. Notably, an emerging pattern across tiers 1–3 is that increasing model complexity does not uniformly improve performance: Motamedi et al. [29] found simpler ARIMA models comparable to multivariate ML approaches for platelet forecasting, echoing the diminishing returns of complexity observed in tier 2. This finding becomes even more pronounced at the regional scale, as discussed in tier 4. Furthermore, facility-level predictions both aggregate individual patient-level demand (tier 1) and directly feed into the regional supply planning discussed in tier 4, positioning tier 3 as the operational bridge between clinical and logistical decision-making.

Tier 4: Regional/National Supply Chain

Goal: Secure Long-Term Blood Supply at Population Level through Strategic Planning

Five studies examined regional forecasting for blood supply planning. Kwon et al. [32] analyzed Korean national health insurance big data using ensemble boosting methods, achieving R^2 up to 0.85 (specifically O+ platelet concentrates) for monthly platelet demand prediction. The model incorporated demographic factors, seasonal patterns, and healthcare utilization trends. Sarvestani et al. [8] compared ARIMA and ANN for monthly demand forecasting in Shiraz, Iran, finding ARIMA superior for the majority of blood groups in capturing aggregated temporal trends using MAE and MSE metrics. Liu et al. [33] applied Seasonal ARIMA (SARIMA) to Chinese hematology department red blood cell demand, achieving MAPE of 13.69% for monthly forecasting. Shih and Rajendran [34] compared seven time series methods and ML algorithms for forecasting Taiwan Blood Services Foundation's RBC supply, finding Seasonal ARIMA and Seasonal Exponential Smoothing outperformed neural network approaches. Shokouhifar and Ranjbarimesan [35] applied multivariate LSTM incorporating COVID-19 case data for blood donation and demand forecasting at

the Tehran Blood Center, reporting correlation coefficients of 0.98 (donations) and 0.97 (demand) for weekly prediction. These notably high correlations should be interpreted with caution as the model was trained on only 85 weeks (20 months) of data, and deep learning models, particularly LSTMs, are known to be susceptible to overfitting on small datasets, potentially inflating apparent performance.

At regional scales, aggregated trends smooth individual-level noise, making traditional statistical time series models highly effective and often preferable to complex ML approaches [8, 34]. This confirms the pattern observed across tiers 1–3: as the prediction scope broadens from individual patients to populations, the marginal benefit of model complexity diminishes. Platelet demand proved partially predictable due to the scheduled nature of cancer chemotherapy cycles, which constitute a major driver of platelet utilization [32]. The COVID-19 pandemic demonstrated the value of incorporating exogenous variables into demand models as conventional time series methods alone could not capture the abrupt disruptions to both supply and demand [35].

Technical Approaches: Comparison of Methods

Table 2 provides a systematic comparison of the technical approaches employed across the included studies. The choice of method should be guided by the prediction task, available data, and institutional resources.

Key Observations on Method Selection:

- Complexity vs. data availability trade-off: deep learning methods (LSTM) showed superior performance for patient-level prediction when sufficient training data were available [16], but traditional methods (ARIMA) proved equally or more effective for aggregated regional forecasting [8, 33, 34].
- Interpretability considerations: clinical adoption requires model transparency. Studies employing SHAP values [17] or inherently interpretable models [25] facilitate clinician trust and “human-in-the-loop” decision-making.
- Gradient boosting dominance for tabular data: for procedure-related prediction using structured clinical variables, XGBoost [38], LightGBM, and gradient boosting machines consistently achieved strong performance [15, 20, 22], aligning with broader ML benchmarking literature on tree-based methods [39, 40].
- Absence of transformer architectures: notably, none of the included studies employed transformer architectures despite their demonstrated effectiveness in time series forecasting [41]. This likely reflects the relatively small and heterogeneous datasets typical in transfusion medicine, where the data requirements and

Table 2. Comparison of technical approaches for transfusion demand prediction

Method	Advantages	Limitations	Best suited for	Data requirements
ARIMA/SARIMA	Simple implementation; interpretable; robust for aggregated data; captures seasonality; minimal computational resources	Cannot capture nonlinear patterns; assumes stationarity; limited to univariate analysis	Regional forecasting (tier 4); stable demand patterns	≥2 years historical data; regular intervals
Exponential smoothing	Very simple; handles trends and seasonality; fast computation; good baseline	Limited pattern complexity; no exogenous variables; sensitive to outliers	Short-term facility forecasting; resource-limited settings	≥1 year historical data
Random forest/XGBoost	Handles tabular data excellently; feature importance available; robust to outliers; no feature scaling needed	Requires careful feature engineering; limited temporal modeling; may overfit small datasets	Surgical/procedure prediction (tier 2); patient classification	Structured clinical variables; moderate sample sizes
LSTM/deep learning	Captures complex temporal dependencies; learns features automatically; handles multivariate sequences	Requires large datasets; computationally intensive; “black box”; hyperparameter sensitive	Patient trajectories (tier 1); high-frequency data	Large datasets (>10,000); sequential data; GPU resources
Hybrid ML + operations research	Combines prediction with optimization; directly actionable recommendations; considers constraints	Complex implementation; requires OR expertise; model coupling challenges	Inventory optimization (tier 3); ordering decisions	Demand forecasts plus inventory constraints
Ensemble/meta-learners	Often superior performance; reduces variance; robust predictions	Increased complexity; reduced interpretability; higher computational cost	Critical predictions requiring high accuracy	Sufficient data for multiple model training

computational overhead of transformer models may not yet be justified compared to established approaches.

- Hybrid approaches for actionable recommendations: the combination of ML prediction with operations research optimization [27] transforms forecasts into concrete ordering decisions, addressing the “last mile” of clinical implementation.
- Product-specific prediction horizons: as Table 3 illustrates, the short platelet shelf life (5–7 days) necessitates 3–7 day forecasting windows, whereas the longer viability of pRBCs (35–42 days) permits weekly to monthly horizons. Platelet demand is partially predictable through chemotherapy schedules and hematology-oncology patient trajectories [18, 29], and even modest improvements in platelet forecasting translate to substantial cost savings given high unit costs and wastage rates of 5–20% [5].

Risk-of-Bias Assessment

Table 4 summarizes PROBAST assessments for all included studies. Participant selection, predictor definition, and outcome determination were generally rated as low risk across studies. The analysis domain was the primary source of bias concerns: only 6 of 24 studies achieved low risk in this domain. External validation was performed by only two studies: Flint et al. [25] used

temporal and multi-site validation across 37 centers, and Jo et al. [22] validated on an independent institution. Rafiei et al. [16] and Li et al. [27] employed year-wise temporal validation at single institutions, which provides stronger evidence than cross-validation but does not test geographic generalizability. Common analysis concerns included reliance on internal validation only (18 studies), use of k-fold cross-validation on time series data risking temporal leakage, univariate feature pre-selection, and small sample sizes. One study (Sarvestani et al. [8]) was rated high risk due to acknowledged model overfitting on only 96 monthly observations.

Discussion

Summary of Evidence

This systematic review demonstrates that AI and ML applications have shown promising performance for transfusion demand prediction across all four tiers of our framework. Patient-individual models achieved AUCs up to 0.97 for short-term prediction, procedure-related models showed AUCs of 0.70–0.91 depending on surgical complexity, facility-level optimization approximately halved platelet outdated rates and achieved MAPEs as low as 4.18%, and regional forecasting achieved R^2 values up to 0.85.

Table 3. Prediction horizons by blood product type

Blood product	Shelf life	Optimal prediction horizon	Key demand drivers
Platelet concentrates (thrombocyte, TK)	5–7 days	3–7 days	Chemotherapy schedules; hematologic malignancies; cardiac surgery
Packed red blood cells (erythrocyte, EK)	35–42 days	Weekly to monthly	Surgical volumes; trauma patterns; chronic anemia management

However, whether these metrics constitute “clinically meaningful” performance depends on tier-specific operational requirements that differ substantially across contexts. For tiers 1 and 2 (individual/procedure-level prediction), clinical meaningfulness must account for outcome prevalence and the asymmetric costs of errors: a false negative (failing to prepare blood for a patient who needs it) carries far greater clinical risk than a false positive (unnecessary cross-matching). In low-prevalence settings, even models with AUCs above 0.90 may generate substantial numbers of false positives. For tier 2, the relevant benchmark is whether the model improves upon current MSBOS guidelines. Feng et al.’s [20] finding that LightGBM outperformed clinician judgment for 85% of cases provides this comparison, but most studies lack such a baseline. For tier 3, acceptable MAPE must be evaluated relative to buffer stock levels and product shelf life: an MAPE of 18% for platelets with a 5-day shelf life has fundamentally different implications than the same error for pRBCs with 42 days of viability. Institutions with larger platelet inventories can absorb greater forecast error, making deployment thresholds inherently institution specific. For tier 4, acceptable error margins should be calibrated against the planning horizon and the ability to adjust supply through donor recruitment. The field would benefit from establishing tier-specific performance benchmarks that incorporate these contextual factors, rather than relying on absolute metric thresholds.

The four-tier framework provides a structured approach for matching prediction methods to clinical objectives. Importantly, the goals at each tier are distinct: improving individual care quality (tier 1) requires different approaches than securing population-level supply (tier 4). This distinction should guide both research priorities and implementation strategies.

Cross-Cutting Findings

Three overarching patterns emerge from the synthesis across all four tiers.

Model Complexity versus Performance

A consistent finding is that increasing model complexity yields diminishing returns as the prediction scope broadens. At the individual patient level (tier 1), deep

learning models that capture temporal dependencies in sequential laboratory data achieved the highest performance [16, 18]. For structured clinical variables in tier 2, gradient boosting methods, which are moderately complex but interpretable, proved sufficient. At the facility and regional levels (tiers 3–4), simpler ARIMA models achieved performance comparable to or exceeding more complex approaches [8, 29, 34]. This inverse relationship between aggregation level and required model complexity has practical implications: institutions need not invest in deep learning infrastructure for supply chain forecasting when well-tuned statistical models perform equivalently.

External Validation Gap

Arguably, the most important finding across the included studies is the near-universal reliance on internal validation. Only 4 of 24 studies employed external, multi-site, or prospective validation [22, 24, 25, 27]. Where external validation was performed, performance typically declined: Flint et al.’s [25] ACSePT tool achieved ROC of 0.70 in external validation, notably lower than the internally validated AUCs exceeding 0.85 reported in several other tier 2 studies. This pattern suggests that reported performance metrics across the field likely overestimate real-world generalizability, and no tier can currently be considered to have sufficient evidence for widespread clinical deployment without further validation.

Readiness for Clinical Deployment

The evidence suggests that tier 3 (facility-level inventory optimization) is closest to deployment readiness as these models address operational rather than individual clinical decisions, reducing the regulatory burden, and have demonstrated concrete cost savings in simulation [11, 27, 30]. However, it is important to note that much of this evidence, including the platelet outdating reductions reported by Guan et al. [11] and Schilling et al. [30], is based on retrospective simulation rather than prospective implementation. Tier 2 procedure-specific models with external validation [22, 25] are also candidates for deployment as decision-support tools for MSBOS modernization. By contrast, tier 1 individual

prediction and tier 4 regional forecasting face greater barriers: the former due to the validation gap and regulatory requirements for clinical decision support, the latter due to the challenge of incorporating unpredictable exogenous shocks (e.g., pandemics, policy changes).

Technical Approach Selection

Our analysis reveals that no single technical approach dominates across all prediction tasks. Rather, method selection should be guided by specific characteristics of the prediction problem:

For Patient-Individual Prediction (Tier 1)
Deep learning approaches, particularly LSTM networks, excel when sufficient sequential patient data are available. The ability to learn temporal patterns in laboratory trajectories (e.g., declining platelet counts during chemotherapy) provides substantial predictive value [18]. However, ensemble methods combining multiple classifiers can achieve comparable performance while offering greater robustness [16].

For Procedure-Related Prediction (Tier 2)
Gradient boosting methods (XGBoost, LightGBM, GBM), logistic regression, and support vector machines consistently performed well for structured clinical data. Feature importance analyses consistently identified preoperative hemoglobin as the dominant predictor, suggesting that simpler models incorporating this key variable may be sufficient for many applications [20, 22, 23].

For Facility-Level Optimization (Tier 3)
Hybrid approaches combining ML forecasting with operations research optimization offer the most actionable outputs. Pure forecasting provides demand estimates, but integration with inventory optimization models generates specific ordering recommendations [27].

For Regional Forecasting (Tier 4)
Traditional time series methods (ARIMA, SARIMA) proved highly effective. At aggregated scales, individual-level noise cancels out, revealing clear temporal patterns amenable to classical statistical approaches [8, 33]. The simplicity and interpretability of these methods are advantages for strategic planning applications.

Limitations of Current Evidence
Despite promising performance metrics, significant limitations constrain clinical implementation.

External Validation Gap
The predominant reliance on internal validation across studies (Table 4) raises concerns about generalizability. Healthcare systems differ substantially in

Table 4. PROBABST risk of bias summary for included studies

Study	Analysis	Overall risk
Chen H. et al. [21]	Unclear*†	Low/moderate
Chen X. et al. [23]	Unclear*¶	Unclear
Engelke et al. [18] (2023a)	Unclear*	Low/moderate
Engelke et al. [19] (2023b)	Unclear*	Low/moderate
Fanoodi et al. [31]	Unclear*	Low/moderate
Feng et al. [20]	Unclear*	Low/moderate
Flint et al. [25]	Low	Low
Guan et al. [11]	Unclear*	Low/moderate
Jo et al. [22]	Low¶	Low
Kwon et al. [32]	Unclear*	Low/moderate
Lang et al. [17]	Unclear*	Unclear
Li et al. [27]	Low	Low
Liu et al. (liver Tx) [24]	Low‡	Low
Liu et al. (regional) [33]	Unclear‡*	Low/moderate
Mitterecker et al. [15]	Unclear*	Low/moderate
Motamedi et al. [29] (2024)	Unclear*	Low/moderate
Motamedi et al. [12] (2025)	Unclear*†	Low/moderate
Rafiei et al. [16]	Unclear*	Low/moderate
Sarvestani et al. [8]	High#	High
Schilling et al. [30]	Low	Low
Shih & Rajendran [34]	Unclear*	Low/moderate
Shokouhifar et al. [35]	Unclear*	Low/moderate
Turkulainen et al. [28]	Low	Low
Wang et al. [26]	Unclear*	Low/moderate

Participants rated low for all studies. Predictors rated low except Chen X. et al. and Lang et al. (unclear). Outcome rated low except Jo et al. and Liu et al. (liver Tx) (unclear). Detailed assessments in online supplementary material. *Internal validation only; †simulation based; ‡prospective validation; ¶univariate screening, no temporal split; ¶external validation; #very small sample (96 observations), acknowledged overfitting.

patient populations, clinical practices, and data infrastructure. Only two studies performed external or multi-site validation: Flint et al. [25] validated ACSePT across 37 centers on 12,529 independent patients and Jo et al. [22] validated on an independent institution. Two additional studies (Rafiei et al. [16] and Li et al. [27]) used year-wise temporal validation at single institutions, testing on unseen years but not on independent populations. Flint et al.’s [25] modest ROC (0.70) in external validation, compared with higher-performing but internally validated models, illustrates the typical performance gap when moving beyond development data.

Heterogeneous Outcome Definitions

Studies variously defined transfusion need as any transfusion, transfusion above threshold volumes, or specific product types. Similarly, prediction horizons ranged from 24 h to 1 month, making direct performance comparisons challenging.

Data Leakage Risk

In retrospective studies, predictors measured temporally close to outcomes may inadvertently encode information about the outcome itself, inflating apparent performance. Careful temporal separation of predictors and outcomes is essential but inconsistently reported.

Publication Bias

Studies reporting positive results are more likely to be published, potentially overstating the achievable performance of AI methods.

Implementation Considerations

A substantial disconnect exists between research demonstrations and clinical deployment. Several barriers must be addressed:

EHR Integration

Models requiring high-frequency laboratory values may be impractical in settings with less frequent monitoring. Real-time data pipelines are technically challenging to implement and maintain.

Data Interoperability

Deploying prediction models across institutions requires standardized data formats and terminologies. Adoption of interoperability standards such as HL7 FHIR and common data models (e.g., OMOP CDM) would facilitate multi-site model deployment and external validation, reducing the need for institution-specific feature engineering and re-training.

Regulatory Pathways

AI-based tools that directly inform individual patient care decisions may require regulatory approval as clinical decision support software (FDA, CE marking). In contrast, facility-level demand forecasting and inventory optimization tools primarily serve operational rather than clinical decision-making and may fall outside traditional CDS regulatory frameworks, though the regulatory landscape continues to evolve.

Clinical Workflow Integration

Predictions must reach decision-makers at actionable timepoints. A 3-day platelet forecast is useful only if it informs collection scheduling or inventory ordering.

Model Maintenance

ML models require ongoing monitoring for performance degradation as patient populations and clinical practices evolve (“model drift”).

Future Directions

Several research priorities emerge from this review. Most critically, multicenter external validation must become a standard requirement for prediction model development in this field. The current predominance of internally validated models limits confidence in generalizability as healthcare systems differ substantially in patient demographics, clinical protocols, transfusion thresholds, and data infrastructure. Prospective validation across diverse settings, including both academic medical centers and community hospitals, is essential before any widespread clinical deployment can be responsibly pursued.

Closely related is the need for standardized outcome definitions across the field. Studies currently vary in how they define transfusion need, ranging from any transfusion event to product-specific thresholds, and employ prediction horizons spanning 24 h to several months. Establishing consensus on transfusion prediction endpoints and reporting standards would not only facilitate meaningful cross-study comparisons but also enable future meta-analyses that could more rigorously quantify the added value of AI-based approaches over conventional methods.

Beyond methodological standardization, a key opportunity lies in the integration of hospital-level prediction models with regional blood center operations. Currently, most models operate in isolation at individual institutions. Connecting facility-level demand forecasts with blood center collection scheduling and inter-hospital distribution networks could optimize the entire supply chain, enabling proactive donor recruitment during anticipated demand surges and more efficient redistribution of products approaching expiration.

The development of explainable AI methods represents another important frontier. Clinical adoption fundamentally depends on clinician trust, which in turn requires model transparency. Continued advancement of inherently interpretable architectures, such as the Temporal Fusion Transformer [41] that offers built-in attention-based interpretability for multi-horizon forecasting, alongside post hoc explanation methods including SHAP, LIME, and layer-wise relevance propagation [42], will be essential for enabling meaningful “human-in-the-loop” decision-making in transfusion medicine.

Federated learning represents a particularly promising avenue for addressing the external validation gap identified in this review. By enabling collaborative model training across institutions without sharing patient data,

federated approaches could produce models that generalize across healthcare systems while preserving data privacy. Recent work has demonstrated the feasibility of federated learning for transfusion prediction in ICU settings [43] and explored privacy-preserving data access frameworks for transfusion medicine [44], though no federated demand forecasting models meeting our inclusion criteria have been published to date. Given that the lack of external validation is the most critical limitation of the current evidence base, federated approaches offer a practical path forward that circumvents the institutional barriers to multi-site data sharing.

Reinforcement learning for inventory management represents another underexplored methodology in this field. While our review identified studies using supervised learning and time series methods for demand forecasting, reinforcement learning, which optimizes sequential decision-making under uncertainty, is naturally suited to the platelet inventory problem, where ordering decisions must balance shortage risk against wastage under stochastic demand and short shelf lives. Although no qualifying studies were identified in our systematic search, applications of reinforcement learning to perishable inventory management in adjacent domains suggest significant potential for blood product inventory optimization.

Blood group-specific demand forecasting also warrants further investigation. While Kwon et al. [32] demonstrated that prediction accuracy varies by blood group (achieving R^2 of 0.85 specifically for O+ platelet concentrates) and Sarvestani et al. [8] reported group-specific differences in ARIMA forecasting accuracy, no study has systematically examined whether predictability differs across blood groups. This has direct operational relevance: O-negative red blood cells serve as the universal donor product and are disproportionately consumed in emergency settings, yet the stochastic nature of emergency demand may make O-negative forecasting inherently more challenging than predicting demand for group-specific products driven by scheduled procedures.

Finally, the emergence of artificial blood products introduces a novel dimension to demand prediction research. As hemoglobin-based oxygen carriers and other blood substitutes advance toward clinical use [45, 46], the computational infrastructure and forecasting methodologies developed for biological blood products may prove foundational for managing these manufactured alternatives. Unlike donor-derived products, artificial substitutes will introduce different supply dynamics, potentially more predictable manufacturing schedules but new challenges related to production capacity planning and cost optimization, requiring adaptation of existing forecasting frameworks.

Conclusion

This systematic review demonstrates that AI applications in blood transfusion demand prediction have achieved clinically meaningful performance across all four tiers of the proposed framework. At the patient-individual level, deep learning and ensemble methods achieved AUCs up to 0.97, enabling proactive transfusion management that can improve individual care quality. Procedure-related models attained AUCs of 0.70–0.91 and offer a data-driven foundation for modernizing surgical blood ordering schedules. At the facility level, ML-based inventory optimization approximately halved platelet outdating rates and substantially reduced ordering frequency, translating directly into cost savings and reduced wastage. Regional forecasting models achieved R^2 values up to 0.85, supporting strategic supply planning at population scale. These results collectively indicate that AI can meaningfully contribute across the entire blood supply chain, from bedside decisions to national logistics.

The choice of modeling approach should be guided by the specific prediction scope, available data infrastructure, and clinical objectives. Deep learning methods such as LSTM networks excel for patient-level temporal prediction when sufficient sequential data are available, while gradient boosting methods are well suited for procedure-related tasks relying on structured clinical variables. At higher levels of aggregation, traditional time series methods including ARIMA and SARIMA remain effective as individual-level noise is smoothed and clear temporal patterns dominate. For inventory optimization, hybrid approaches combining ML forecasting with operations research offer the most actionable outputs by translating demand estimates into concrete ordering decisions. Prediction horizons should be tailored to product characteristics: platelet forecasting should target 3–7 day windows given the short shelf life, whereas red blood cell prediction can extend to weekly or monthly scales.

Despite these promising results, the predominance of internal validation and heterogeneous outcome definitions across the included studies highlight the critical need for standardized, multicenter external validation before widespread clinical deployment. Ultimately, the greatest potential may lie not in any single tier but in the integration of prediction models across tiers, connecting patient-level forecasts with facility inventory systems and regional supply networks to create a coordinated, data-driven blood supply chain. As demographic shifts drive declining donor populations and increasing transfusion demand, such integrated AI-driven approaches offer substantial potential to simultaneously improve patient outcomes and optimize resource utilization.

Acknowledgments

We acknowledge support by the Open Access Publication Fund of the University of Duisburg-Essen. Artificial intelligence tools were used to improve grammar and language clarity in the manuscript. A generative AI tool was also used to assist in creating the visual abstract. All AI-generated content was reviewed and verified by the authors, who retain full responsibility for the accuracy and integrity of the work.

Statement of Ethics

Ethics approval was not required as this systematic review exclusively analyzed previously published literature and did not involve human participants, animal subjects, or identifiable patient data. This review was retrospectively registered on the Open Science Framework (OSF) on February 7, 2026 (DOI: <https://doi.org/10.17605/OSF.IO/YF5ZP>).

Conflict of Interest Statement

M.E. has published primary research in this field (Engelke et al. 2023a, 2023b). Peter Alexander Horn was a member of the journal's Editorial Board at the time of submission. The other authors have no conflict of interest to declare.

References

- 1 Müller MM, Geisen C, Zacharowski K, Tonn T, Seifried E. Transfusion of packed red blood cells: indications, triggers and adverse events. *Dtsch Arztebl Int.* 2015;112:507–18. <https://doi.org/10.3238/arztebl.2015.0507>
- 2 Paglia G, Palsson BØ, Sigurjonsson OE. Systems biology of stored blood cells: can it help to extend the expiration date? *J Proteonomics.* 2012;76:163–7. <https://doi.org/10.1016/j.jprot.2012.08.014>
- 3 Roberts N, James S, Delaney M, Fitzmaurice C. The global need and availability of blood products: a modelling study. *Lancet Haematol.* 2019;6(12):e606–15. [https://doi.org/10.1016/S2352-3026\(19\)30200-5](https://doi.org/10.1016/S2352-3026(19)30200-5)
- 4 Tancred T, Bates I. Improving blood transfusion services. *Best Pract Res Clin Obstet Gynaecol.* 2019;61:130–42. <https://doi.org/10.1016/j.bpobgyn.2019.05.007>
- 5 Jones JM, Sapiano MRP, Mowla S, Bota D, Berger JJ, Basavaraju SV. Has the trend of declining blood transfusions in the United States ended? Findings of the 2019 National Blood Collection and Utilization Survey. *Transfusion.* 2021;61(S2):S1–10. <https://doi.org/10.1111/trf.16449>
- 6 Ng MSY, David M, Middelburg RA, Ng ASY, Suen JY, Tung J-P, et al. Transfusion of packed red blood cells at the end of shelf life is associated with increased risk of mortality—a pooled patient data analysis of 16 observational trials. *Haematologica.* 2018; 103(9):1542–8. <https://doi.org/10.3324/haematol.2018.191932>
- 7 Greinacher A, Fendrich K, Hoffmann W. Demographic changes: the impact for safe blood supply. *Transfus Med Hemother.* 2010;37(3):141–8. <https://doi.org/10.1159/000313949>
- 8 Sarvestani SE, Hatam N, Seif M, Kasraian L, Lari FS, Bayati M. Forecasting blood demand for different blood groups in Shiraz using auto regressive integrated moving average (ARIMA) and artificial neural network (ANN) and a hybrid approaches. *Sci Rep.* 2022;12(1):22031. <https://doi.org/10.1038/s41598-022-26461-y>
- 9 Li N, Pham T, Cheng C, McElfresh DC, Metcalf RA, Russell WA, et al. Blood demand forecasting and supply management: an analytical assessment of key studies utilizing novel computational techniques. *Transfus Med Rev.* 2023;37(4):150768. <https://doi.org/10.1016/j.tmr.2023.150768>
- 10 Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* 2021;372:n71. <https://doi.org/10.1136/bmj.n71>
- 11 Guan L, Tian X, Gombar S, Zemek AJ, Krishnan G, Scott R, et al. Big data modeling to predict platelet usage and minimize wastage in a tertiary care system. *Proc Natl Acad Sci USA.* 2017;114(43):11368–73. <https://doi.org/10.1073/pnas.1714097114>
- 12 Motamedi M, Dawson J, Li N, Down D. Blood platelet inventory management: incorporating data-driven demand forecasts. *Health Care Manag Sci.* 2025;28(2):191–206. <https://doi.org/10.1007/s10729-025-09706-w>
- 13 Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PRO-BAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med.* 2019;170(1):51–8. <https://doi.org/10.7326/M18-1376>
- 14 Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ.* 2008;336(7650):924–6. <https://doi.org/10.1136/bmj.39489.470347.AD>
- 15 Mitterecker A, Hofmann A, Trentino KM, Lloyd A, Leahy MF, Schwarzbauer K, et al. Machine learning-based prediction of transfusion. *Transfusion.* 2020;60(9):1977–86. <https://doi.org/10.1111/trf.15935>
- 16 Rafiei A, Moore R, Choudhary T, Marshall C, Smith G, Roback JD, et al. Robust meta-model for predicting the likelihood of receiving blood transfusion in non-traumatic intensive care unit patients. *Health Data Sci.* 2024;4:0197. <https://doi.org/10.34133/hds.0197>
- 17 Lang FF, Liu LY, Wang SW, Yang Z-Y, Chen X, Xu S-D. Predictive modeling of perioperative blood transfusion in lumbar posterior interbody fusion using machine learning. *Front Physiol.* 2023;14:1306453. <https://doi.org/10.3389/fphys.2023.1306453>
- 18 Engelke M, Schmidt CS, Baldini G, Parmar V, Hosch R, Borys K, et al. Optimizing platelet transfusion through a personalized deep learning risk assessment system for demand management. *Blood.* 2023;142: 2315–26. <https://doi.org/10.1182/blood.2023021172>

Funding Sources

The ReMeDi:Blut (“Register für medizinische Daten und Antigeneigenschaften von Blutprodukten”) project is supported by the Federal Ministry of Research, Technology and Space of Germany (BMFTR) as part of the funding initiative “Medical Technology Solutions for Digital Healthcare,” funding reference No. 13GW0620A granted to PAH.

Author Contributions

M.E.: literature search, data extraction, and manuscript drafting. O.A.C.-K.: literature screening and data extraction. J.K.: critical revision and supervision. P.A.H.: conceptualization and critical revision. All authors discussed the results and contributed to the manuscript, including approval of the final manuscript.

Data Availability Statement

All data generated or analyzed during this study are included in this article and its supplementary material files. Further inquiries can be directed to the corresponding author.

- 19 Engelke M, Brieske CM, Parmar V, Flaschel N, Kureishi A, Hosch R, et al. Predicting individual patient platelet demand in a large tertiary care hospital using machine learning. *Transfus Med Hemother*. 2023;50(4):277–85. <https://doi.org/10.1159/000528428>
- 20 Feng Y, Xu Z, Sun X, Wang D, Yu Y. Machine learning for predicting preoperative red blood cell demand. *Transfus Med*. 2023;31(4):262–70. <https://doi.org/10.1111/tme.12794>
- 21 Chen H, Cao B, Yang J, Ren H, Xia X, Zhang X, et al. Construction and effect evaluation of prediction model for red blood cell transfusion requirement in cesarean section based on artificial intelligence. *BMC Med Inform Decis Mak*. 2023;23(1):213. <https://doi.org/10.1186/s12911-023-02286-1>
- 22 Jo C, Ko S, Shin WC, Han H-S, Lee MC, Ko T, et al. Transfusion after total knee arthroplasty can be predicted using the machine learning algorithm. *Knee Surg Sports Traumatol Arthrosc*. 2020;28(6):1757–64. <https://doi.org/10.1007/s00167-019-05602-3>
- 23 Chen X, Pan J, Li Y, Tang R, Wu B, Wei P. Application of machine learning model in predicting the likelihood of blood transfusion after hip fracture surgery. *Aging Clin Exp Res*. 2023;35(11):2643–56. <https://doi.org/10.1007/s40520-023-02550-4>
- 24 Liu L-P, Zhao Q-Y, Wu J, Luo Y-W, Dong H, Chen Z-W, et al. Machine learning for the prediction of red blood cell transfusion in patients during or after liver transplantation surgery. *Front Med*. 2021;8:632210. <https://doi.org/10.3389/fmed.2021.632210>
- 25 Flint AWJ, Bailey M, Reid CM, Smith JA, Tran L, Wood EM, et al. Preoperative identification of cardiac surgery patients at risk of receiving a platelet transfusion: the Australian Cardiac Surgery Platelet Transfusion (ACSePT) risk prediction tool. *Transfusion*. 2020;60(10):2272–83. <https://doi.org/10.1111/trf.15990>
- 26 Wang Y, Zhang W, Rao Q, Ma Y, Ding X, Zhang X, et al. Forecasting demands of blood components based on prediction models. *Transfus Clin Biol*. 2024;31(3):141–8. <https://doi.org/10.1016/j.traci.2024.04.003>
- 27 Li N, Arnold DM, Down DG, Barty R, Blake J, Chiang F, et al. From demand forecasting to inventory ordering decisions for red blood cells through integrating machine learning, statistical modeling, and inventory optimization. *Transfusion*. 2022;62(1):87–99. <https://doi.org/10.1111/trf.16739>
- 28 Turkulainen EV, Wemelsfelder ML, Janssen MP, Arvas M. A robust autonomous method for blood demand forecasting. *Transfusion*. 2022;62(6):1261–8. <https://doi.org/10.1111/trf.16870>
- 29 Motamedi M, Dawson J, Li N, Down DG, Heddle NM. Demand forecasting for platelet usage: from univariate time series to multivariable models. *PLoS One*. 2024;19(4):e0297391. <https://doi.org/10.1371/journal.pone.0297391>
- 30 Schilling M, Rickmann L, Hutschenreuter G, Spreckelsen C. Reduction of platelet outdating and shortage by forecasting demand with statistical learning and deep neural networks: modeling study. *JMIR Med Inform*. 2022;10(2):e29978. <https://doi.org/10.2196/29978>
- 31 Fanoodi B, Malmir B, Jahantigh FF. Reducing demand uncertainty in the platelet supply chain through artificial neural networks and ARIMA models. *Comput Biol Med*. 2019;113:103415. <https://doi.org/10.1016/j.combiomed.2019.103415>
- 32 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. *Digit Health*. 2024;10:20552076231224245. <https://doi.org/10.1177/20552076231224245>
- 33 Liu J, Yuan Q, Cai L, Cai Z, Xu X. Time series forecasting of red blood cell demand in hematology patients using SARIMA and exponential smoothing models: a retrospective analysis in a Chinese tertiary hospital. *Front Med*. 2025;12:1582277. <https://doi.org/10.3389/fmed.2025.1582277>
- 34 Shih H, Rajendran S. Comparison of time series methods and machine learning algorithms for forecasting Taiwan Blood Services Foundation's blood supply. *J Healthc Eng*. 2019;2019:1–6. <https://doi.org/10.1155/2019/6123745>
- 35 Shokouhifar M, Ranjbarimesan M. Multivariate time-series blood donation/demand forecasting for resilient supply chain management during COVID-19 pandemic. *Clean Logist Supply Chain*. 2022;5:100078. <https://doi.org/10.1016/j.clscn.2022.100078>
- 36 Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997;9(8):1735–80. <https://doi.org/10.1162/neco.1997.9.8.1735>
- 37 Carson JL, Guyatt G, Heddle NM, Grossman BJ, Cohn CS, Fung MK, et al. Clinical practice guidelines from the AABB: red blood cell transfusion thresholds and storage. *JAMA*. 2016;316(19):2025–35. <https://doi.org/10.1001/jama.2016.9185>
- 38 Chen T, Guestrin C. XGBoost: a scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*; 2016. p. 785–94. <https://doi.org/10.1145/2939672.2939785>
- 39 Breiman L. Random forests. *Mach Learn*. 2001;45(1):5–32. <https://doi.org/10.1023/A:1010933404324>
- 40 Grinsztajn L, Oyallon E, Varoquaux G. Why do tree-based models still outperform deep learning on typical tabular data? *Adv Neural Inf Process Syst*. 2022;35:507–20.
- 41 Lim B, Arık SÖ, Loeff N, Pfister T. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *Int J Forecast*. 2021;37(4):1748–64. <https://doi.org/10.1016/j.ijforecast.2021.03.012>
- 42 Keyl J, Keyl P, Montavon G, Hosch R, Brehmer A, Mochmann L, et al. Decoding pan-cancer treatment outcomes using multimodal real-world data and explainable artificial intelligence. *Nat Cancer*. 2025;6(2):307–22. <https://doi.org/10.1038/s43018-024-00891-1>
- 43 Choudhury A. A federated learning model for the prediction of blood transfusion in intensive care units. *Stud Health Technol Inform*. 2025;322:640–4. <https://doi.org/10.3233/SHIT250326>
- 44 Cohn CS, Knudson CM, Engelbrecht-Wiggans ER, Stubbs JR, Harm SK, Raval JS. Privacy-preserving federated data access and federated learning: improved data sharing and AI model development in transfusion medicine. *Transfusion*. 2025;65:125–33. <https://doi.org/10.1111/trf.18098>
- 45 Spahn DR, Goodnough LT. Alternatives to blood transfusion. *Lancet*. 2013;381(9880):1855–65. [https://doi.org/10.1016/S0140-6736\(13\)60808-9](https://doi.org/10.1016/S0140-6736(13)60808-9)
- 46 Savla C, Palmer AF. Hemoglobin-based oxygen carriers: where are we now in 2023? *Medicina*. 2023;59:396. <https://doi.org/10.3390/medicina59020396>