

Machine Learning for Donor and Resource Management: A Systematic Review Applied to Thalassemia Care

Vedant Banaitkar¹, Dipak Dhumma², Vynktesh Mande³, Prathamesh Kulkarni⁴, Ravindra Mule⁵

^{1,2,3,4}Dept. of Artificial Intelligence and Data Science, PVG's College of Engineering and Technology, Pune, India.

⁵Faculty Guide, Dept. of AI & DS, PVG's COET, Pune, India.

Email ID: vedantbanaitkar@gmail.com¹

Abstract

The management of Thalassemia depends on the timely and continuous availability of compatible blood transfusions, which places ongoing coordination demands on patients, donors, and blood banks. Machine learning (ML) has increasingly been applied to healthcare logistics in recent years, particularly for studying donor behavior, forecasting blood inventory, and supporting operational decision-making. However, these efforts are largely oriented toward system-level efficiency, with limited focus on transfusion-dependent patient care. This paper reviews existing studies that apply ML techniques to blood donor prediction, inventory management, and donor retention. The surveyed work shows that ML-based approaches can improve population scale planning and resource utilization. At the same time, most solutions are designed for aggregate optimization and do not directly address the continuity required in chronic transfusion settings. For individuals with Thalassemia, care depends less on short-term efficiency and more on the sustained availability of suitable donors over long periods. Based on the reviewed literature, there is a need for ML frameworks that make use of longitudinal data and explicitly support patient-linked donor coordination in long-term transfusion management.

Keywords: machine learning, systematic review, thalassemia, blood transfusion, donor management, predictive modeling, healthcare logistics, patient-centered analytics.

1. Introduction

Thalassemia is a hereditary blood disorder characterized by impaired hemoglobin production, and individuals diagnosed with it often depend on lifelong, periodic transfusions to maintain stable health. Unlike emergency transfusion cases, where demand is unpredictable and episodic, Thalassemia introduces a steady and recurring requirement for compatible blood. This ongoing dependency creates not only medical complexity but also sustained logistical pressure on patients, donors, and healthcare institutions. Because transfusions typically occur at regular intervals, effective management extends beyond simply sourcing blood units; it involves maintaining reliable donor compatibility and ensuring continuity of care over extended periods. Any disruption in this chain may delay treatment, potentially leading to complications such as iron overload or organ dysfunction, while simultaneously imposing emotional and financial strain on affected

families. In parallel, healthcare systems over the past decade have increasingly adopted data-driven approaches to improve operational efficiency. Machine learning (ML), in particular, has gained prominence as a tool for extracting actionable insights from large clinical and logistical datasets. Its applications span appointment scheduling, demand forecasting, and supply chain optimization, among others. Within blood donation systems, ML techniques have been used to analyze donor participation patterns, predict retention, and improve recruitment strategies, thereby supporting more stable inventory management. Despite these advances, the majority of ML-driven solutions emphasize system-wide efficiency rather than individualized continuity of care. Most studies focus on aggregate donor populations or large-scale logistics, leaving the long-term coordination needs of transfusion-dependent patients relatively unexplored.

For conditions such as Thalassemia, however, the central challenge is not merely maximizing overall supply but ensuring the sustained availability of suitable donors for specific individuals over time. This review therefore examines existing research on ML applications in donor prediction, retention analysis, and blood resource management with particular attention to their relevance for chronic transfusion care. By synthesizing recent contributions, the study seeks to clarify prevailing methodological trends and to highlight opportunities for developing patient-linked, longitudinal ML frameworks tailored to Thalassemia management.

2. Review Methodology

To understand how machine learning (ML) has been applied within blood donor and resource management, this review followed a structured yet flexible screening process. Rather than merely cataloguing prior work, the intention was to examine methodological tendencies, identify recurring technical approaches, and observe where gaps persist particularly in relation to transfusion-dependent care.

a) Search Strategy:

The literature search was conducted across widely recognized academic databases, including IEEE Xplore, ACM Digital Library, Springer Link, and Google Scholar. These platforms were selected to ensure adequate representation of both engineering-oriented and healthcare-focused research. A combination of keywords was used in varying arrangements to expand retrieval scope. Primary terms included “machine learning,” “blood donation,” “donor prediction,” “donor retention,” “appointment scheduling,” and “healthcare supply chain.” Boolean operators and related term variations were incorporated where necessary to capture interdisciplinary contributions that might not fall under a single indexing category.

b) Inclusion and Exclusion Criteria:

To maintain clarity and relevance, inclusion and exclusion conditions were defined prior to the screening process. Studies were included if they met the following criteria: (1) publication in English between 2019 and 2025; (2) availability as peer-reviewed journal articles, conference proceedings, or technically detailed preprints; and (3) explicit

implementation or evaluation of computational models particularly ML techniques within blood donation systems or healthcare resource management. Publications that were purely editorial, opinion-based, or lacking methodological transparency were excluded. This filtering ensured that the final corpus emphasized technically substantive contributions rather than conceptual commentary alone.

c) Selection Process:

The initial database queries generated a broad pool of potentially relevant records. Titles and abstracts were first screened to remove clearly unrelated studies. Papers that satisfied preliminary relevance criteria were then examined in full to assess methodological depth and alignment with the defined scope. Following this two-stage evaluation, 19 studies were retained for detailed analysis. For each selected paper, information regarding research objectives, modeling strategies, data sources, reported outcomes, and stated limitations was extracted. These elements were subsequently synthesized to reveal overarching trends and recurring methodological constraints within the field.

3. Analysis of Surveyed Literature

A closer examination of the selected studies reveals three dominant directions in current research: (1) predictive modeling of donor behavior, (2) architectural development of management platforms, and (3) optimization of operational workflows and supply chains. While these themes overlap in practice, they reflect distinct methodological priorities within the field.

a) Predictive Modeling of Donor Behavior:

A substantial portion of the literature concentrates on forecasting donor-related behavior using supervised learning techniques, most commonly classification and regression models. The primary objective across these studies is to improve donor retention and enhance the stability of blood supply systems. For example, Kauten et al. [1] evaluated multiple ML algorithms and reported that Random Forest achieved comparatively strong performance in predicting donor return patterns. Similarly, Salazar-Concha and Ramírez-Correa [2] applied a Decision

Tree model to analyze psychological determinants influencing donation intention. Although these studies approach the problem from different analytical perspectives behavioral versus operational they both attempt to anticipate donor participation before shortages occur. Beyond individual donor prediction, Valero et al. [3] extended predictive modeling to system-level inventory forecasting. Their regression-based framework demonstrated improvements in estimating blood shortages under constrained resource conditions. However, across these efforts, a recurring limitation emerges: model performance is heavily dependent on the representativeness and scope of the training data. Many studies rely on geographically limited datasets or synthetically generated data, raising legitimate concerns about external validity and cross-context generalizability.

b) System Architecture and Management Platforms

Another stream of research focuses less on prediction and more on the structural design of digital management systems. Early contributions by Rahman [4] and Shah et al. [5] emphasized database architecture, proposing relational and object-oriented frameworks to centralize blood bank records. These works primarily address data organization and system efficiency rather than predictive intelligence. Subsequent developments moved toward more integrated platforms. Zulfiqar Ali and Ilyas [6] implemented a web-based blood donation system capable of centralized registration and real-time search during emergencies. Building on this trajectory, Ben Elmir et al. [7] incorporated machine learning directly into a blood management platform to forecast demand and reduce wastage. Taken together, these architectural studies illustrate a gradual transition from static database management toward data-driven, analytics enabled systems. Nonetheless, many implementations remain either conceptual or limited in scope, with advanced functionalities such as intelligent donor matching or adaptive scheduling frequently identified as future work rather than completed components.

c) Optimization of Scheduling and Supply Chains

Operational efficiency represents a third major focus within the surveyed literature. Here, research tends to separate into micro-level and macro-level optimization problems. At the micro level, the emphasis lies on appointment scheduling and patient attendance prediction. Toffaha et al. [8] and Deina et al. [9] developed interpretable ML models aimed at reducing hospital no-shows. Their approaches highlight the growing importance of transparency in predictive healthcare systems, particularly when models directly influence scheduling decisions. At the macro level, studies such as Khalilpourazari et al. [10] address large-scale supply chain optimization, especially under crisis scenarios. These models often incorporate multi-objective optimization techniques to balance cost, demand satisfaction, and distribution efficiency. Broader reviews by Niu et al. [11] and Ala and Chen [12] confirm that heuristic and evolutionary algorithms such as Genetic Algorithms remain prevalent in healthcare scheduling research. However, despite methodological sophistication, many optimization models simplify clinical realities. Factors crucial to transfusion medicine, including donor recipient compatibility constraints and longitudinal patient requirements, are frequently treated as secondary considerations. This simplification limits direct applicability to chronic transfusion-dependent settings.

4. Literature Survey Matrix and Discussion

Table I consolidates the core characteristics of the 19 studies reviewed in this work, summarizing their aims, main outcomes, and reported limitations. The discussion that follows contextualizes these contributions, highlighting shared themes and methodological distinctions among them.

a) A. Discussion of Individual Studies

A closer reading of the selected contributions reveals both methodological diversity and recurring structural limitations across the field. Vassalli et al. [13] examine the socio-economic dimensions of blood donation through a micro-founded model, illustrating how financial incentives interact differently with social trust across European contexts. In contrast, Valero et al. [3] focus on operational constraints, demonstrating through synthetic datasets that heuristic allocation strategies

can significantly improve transfusion request acceptance while enabling shortage forecasting. As shown in Table 1.

Table 1 Literature Survey Matrix

Ref.	Paper Title	Key Finding	Identified Gap / Limitation
[13]	Modeling Blood Donation Behavior and Social Cohesion in Europe.	Financial incentives are not universally effective; they can crowd out altruism in high-trust countries but work in low trust environments.	The analysis is correlational, not causal. Future work should use quasi- experimental methods to test causal prediction.
[3]	Optimizing Blood Transfusions and Predicting Shortages in Resource Constrained Areas.	Heuristic matching algorithms improved blood request acceptance by 47.6%, and Linear Regression was effective for shortage prediction.	The study was based on a synthetic dataset. Real-world data is needed for validation and improvement.
[8]	Predicting Hospital, No Shows: Interpretable Machine Learning Models Approach.	Interpretable ML models like FIGS and TAO Tree achieved high accuracy (~87.5%) in predicting patient no-shows.	The model was trained on a single dataset, which may limit its generalization to other healthcare systems.
[9]	Decision analysis framework for predicting no-shows to Appointments using machine learning algorithms.	A novel framework using Symbolic Regression and Instance Hardness Threshold (IHT) yielded superior sensitivity (≥ 0.94) for no-show prediction.	The predictors were limited by what was available in the dataset; future work needs purpose-built datasets.
[11]	A Review of Optimization Studies for System Appointment Scheduling.	Genetic Algorithms and simulation optimization are the most common methods for appointment scheduling, with deep learning emerging as a core technology.	Concludes that more practical, domain-specific research is needed to tailor optimization algorithms for medical constraints.
[7]	Smart Platform for Data Blood Bank Management: Forecasting Demand in Blood Supply Chain Using Machine Learning.	An integrated smart platform using ML for demand forecasting increased blood collection by 11% and reduced wastage by 20%.	The system is highly dependent on continuous, high quality data. Future work needs to add a data cleaning module.

[14]	Use of Machine Learning in Optimizing Medical Appointment Schedules.	A qualitative review confirming that ML is feasible and effective for optimizing medical appointment scheduling by reducing wait times.	The paper is a high-level review and recommends more specific future studies on predictive scheduling opportunities.
[12]	Appointment Scheduling Problem in Complexity Systems of the Healthcare Services: A Comprehensive Review.	A comprehensive review identifying simulation and AI as the most practical approaches for optimizing appointment scheduling.	A key gap is the need to model and optimize multistage and multi appointment health processes.
[5]	Blood Bank Management and Inventor Control Database Management System.	A well-designed, cloud-hosted relational database can significantly improve the efficiency of blood bank data management.	The paper presents a conceptual design; the GUI-based application to interact with the database is cited as future work.
[15]	A Smart Chatbot for Interactive Management in Beta Thalassemia Patients.	A rule-based chat-bot was found to be a useful and easy-to-use tool for Thalassemia patients to self-manage their assessments.	The Chabot's knowledge base was limited and needs to be expanded to cover all medical complications of the disease.
[16]	A Systematic Review on Smart Blood Bank System: Taxonomy, Motivations, Challenges, Study Directions and Recommendations.	A systematic review that identifies key challenges in smart blood bank systems, including data security, privacy, and connectivity.	The review itself was limited by a lack of access to certain databases, potentially overlooking the latest research.
[2]	Predicting the Intention to Donate Blood among Blood Donors Using a Decision Tree Algorithm.	A Decision Tree model predicted donor re-donation intention with 84.17% accuracy using a minimal set of variables from the Theory of Planned Behaviour.	Results are not generalizable due to a small, culturally-specific sample (Chilean donors).
[1]	Predicting Blood Donors Using Machine Learning Techniques.	Random Forest was the most reliable model (80.1% sensitivity) for predicting donor retention, with scheduling data being the most important feature.	The study used data from a single regional blood center, making comparisons with other regions difficult.
	Recent intelligent Approaches for	A comparative review finding that no single existing system	Concludes the main gap is the need for a single, integrated

[17]	Managing and Optimizing smart Blood Donation process.	integrates all de- sired factors (prediction, optimization, geo- location, etc.).	system that combines all key functionalities.
[4]	Blood Bank Management System	Proposes a conceptual design for a web-based, centralized blood bank management system using Object- Oriented analysis.	The work is a design document; the actual implementation of the proposed system is stated as future scope.
[18]	Patient-Centered Appointment Scheduling: a Call for Autonomy, Continuity, and Creativity	Argues for a patient-centered, decentralized scheduling model managed by the primary care team that knows the patient.	The paper is a perspective piece that explicitly calls for research into applying AI and ML to this scheduling model.
[10]	Designing an efficient blood supply chain network in cri- sis: neural learning, optimization and case study	A multi-objective optimization model successfully designed a blood supply chain for disaster scenarios, reducing unsatisfied demand to zero.	The model did not incorporate blood group compatibility and needs to handle uncertainty using fuzzy math or robust optimization.
[19]	Matching Returning Donors to Projects on Philanthropic Crowdfunding Platforms	A structural econometric model predicted donor choices on a crowd- funding platform more accurately than standard collaborative filtering bench- marks.	The evaluation was a simulation based on historical data; a live field study is needed to validate the approach.
[6]	Design and Development of a Web Platform for Blood Donation Management	A developed web platform provides a centralized solution for donor registration and real- time searches in emergency situations.	Future enhancements like AI- based matching and automated reminders are not part of the current implementation.

Several studies concentrate on appointment-level efficiency. Toffaha et al. [8] and Deina et al. [9] explore interpretable machine learning models for predicting patient no-shows, emphasizing the balance between predictive performance and transparency. Their findings reflect a broader trend toward explainable decision-support systems in healthcare. From a systems perspective, Ben Elmir et al. [7] integrate demand forecasting within a smart blood bank platform, re- porting measurable improvements in collection efficiency and reduced wastage. Complementary reviews by Niu et al. [11]

and Ala and Chen [12] synthesize optimization techniques for appointment scheduling, noting the continued prominence of heuristic and evolutionary methods while highlighting the need for models that better reflect real-world clinical complexity. Infrastructure-focused studies by Rahman [4] and Shah et al. [5] primarily address database architecture and centralized record management. Similarly, Zulfiqar Ali and Ilyas [6] de- scribe a web-based emergency donation platform that enhances real-time coordination, though advanced predictive capabilities remain under development. On the

behavioral modeling side, Salazar-Concha and Ram'irez-Correa [2] employ Decision Trees to estimate re-donation intention within a culturally specific cohort, whereas Kauten et al. [1] identify Random Forest as particularly effective for predicting donor retention in regional datasets. Both studies, however, are constrained by localized data sources. Other contributions expand the scope beyond direct blood bank systems. Alturaiki et al. [15] investigate a rule-based chatbot supporting Thalassemia self-management, while Ismael et al. [16] develop a taxonomy of smart blood bank challenges, including security and interoperability concerns. Khalilpourazari et al. [10] address crisis-oriented supply chain optimization through a multi-objective framework, though compatibility constraints remain simplified. Collectively, these studies demonstrate meaningful technical progress. At the same time, they reveal that most solutions remain either system-centric, dataset-limited, or narrowly scoped leaving individualized, longitudinal donor coordination largely unexplored.

5. Identified Research Gap

When the surveyed studies are considered collectively, a consistent pattern becomes visible. Machine learning has been effectively applied to large-scale donor analytics, inventory forecasting, and supply chain coordination. However, comparatively little attention has been directed toward the sustained, patient-level demands associated with chronic transfusion-dependent conditions such as Thalassemia. Most existing models are constructed using broad, cross-sectional datasets designed to capture aggregate donor behavior or optimize institutional workflows. These approaches are valuable for improving overall system efficiency, yet they are not inherently structured to address individualized continuity of care. In the context of Thalassemia management, the logistical question extends beyond predicting whether donors, in general, will return. Instead, it centers on whether specific compatible donors can be relied upon repeatedly within clearly defined transfusion intervals. This distinction may appear subtle, but it has important methodological implications. Models trained on large heterogeneous datasets often prioritize population-level trends,

whereas chronic transfusion care requires sensitivity to small, stable donor groups and recurring temporal patterns. The perspective advanced by Matulis and McCoy [18], which advocates patient-centered scheduling, aligns conceptually with this need. Nevertheless, practical implementations that integrate such principles with advanced ML techniques remain limited. Accordingly, the principal gap lies not in the absence of predictive modeling itself, but in the absence of models explicitly designed for longitudinal, patient-linked donor coordination. Addressing this gap will require frameworks capable of learning from relatively sparse yet temporally structured data and adapting predictions to individualized transfusion cycles. Without this shift, system-wide optimization efforts risk overlooking the unique requirements of patients who depend on uninterrupted, long-term transfusion support.

6. Future Directions

Building on the identified gap, advancing this research area will require a shift toward more specialized and context-aware modeling strategies. Rather than extending existing population-level frameworks, future investigations may benefit from focusing on approaches that reflect the individualized and recurring nature of transfusion-dependent care. Several directions appear particularly promising:

- **Temporal and Sequential Modeling:** Chronic transfusion management inherently follows a recurring time-line. Models designed to capture sequential dependencies such as Recurrent Neural Networks (RNNs), Long Short-Term Memory networks (LSTMs), or Transformer-based architectures could be adapted to learn donor participation patterns over repeated transfusion cycles. Emphasis should be placed not only on predictive accuracy but also on stability across extended time horizons.
- **Context-Specific Feature Design:** Many existing studies rely heavily on demographic or static operational variables. For transfusion-dependent contexts, additional features may prove relevant, including

transfusion interval regularity, historical donor–recipient compatibility, communication responsiveness, and external temporal factors such as seasonal events or public holidays. Thoughtful feature construction will likely play a central role in improving model relevance at the patient level.

- **Interpretable and Clinically Transparent Systems:** Given the clinical sensitivity of donor selection and scheduling decisions, predictive systems should support interpretability. Explainable AI (XAI) techniques can help clarify the reasoning behind donor recommendations or suggested scheduling adjustments. Such transparency is particularly important in settings where healthcare providers must justify decisions affecting patient safety.
- **Privacy Conscious Collaborative Learning:** The sensitive nature of both patient and donor information necessitates careful handling of data. Approaches such as federated learning or distributed model training offer mechanisms for multi-institutional collaboration without centralized data sharing. Future research should examine how such privacy-aware strategies can be integrated without compromising predictive performance.

7. Conclusion

This review has examined recent scholarly efforts exploring the integration of machine learning within blood donor coordination and resource management systems. The surveyed studies collectively demonstrate meaningful progress in predicting donor participation, designing digital management infrastructures, and optimizing large-scale logistical operations. These advancements indicate that ML has become an established analytical tool within the broader blood supply ecosystem. At the same time, the analysis highlights an imbalance in research focus. While system-wide optimization and population level forecasting have received substantial attention, comparatively limited work addresses the sustained, patient specific demands

associated with chronic transfusion-dependent conditions such as Thalassemia. The distinction between improving overall efficiency and ensuring individualized continuity of care is more than conceptual it has practical implications for how predictive models are structured and evaluated. Moving forward, greater emphasis on longitudinal, patient linked modeling frameworks may help bridge this gap. Approaches that account for recurring transfusion cycles, stable donor compatibility, and adaptive scheduling could contribute to more reliable and personalized transfusion management. Strengthening this research direction has the potential not only to enhance operational stability but also to improve the long- term quality of care for individuals who depend on consistent transfusion support.

References

- [1]. C. Kauten, A. Gupta, X. Qin, and G. G. Richey, “Predicting blood donors using machine learning techniques,” *Information Systems Frontiers*, 2021, published online.
- [2]. C. Salazar-Concha and P. Ramírez-Correa, “Predicting the intention to donate blood among blood donors using a decision tree algorithm,” *Symmetry*, vol. 13, no. 8, p. 1353, 2021.
- [3]. M. Valero, E. A. Belfarsi, and S. Brubaker, “Optimizing blood transfusions and predicting shortages in resource-constrained areas,” in *Proceedings of BIOSTEC, 2025*, data from spreadsheet indicates 2025 and BIOSTEC proceedings.
- [4]. M. M. Rahman, “Blood bank management system,” *ResearchGate*, 2021, project Report.
- [5]. A. Shah et al., “Blood bank management and inventory control database management system,” in *Procedia Computer Science*, vol. 215. Elsevier, 2022, pp. 406–416.
- [6]. F. Zulfiqar Ali and A. Ilyas, “Design and development of a web platform for blood donation management,” Preprint, N/A, year not available in source.
- [7]. W. Ben Elmir et al., “Smart platform for data blood bank management: Forecasting demand in blood supply chain using machine

- learning,” *Information*, vol. 14, no. 11, p. 593, 2023.
- [8]. K. M. Toffaha et al., “Predicting hospital no-shows: Interpretable machine learning models approach,” *IEEE Access*, vol. 12, pp. 18 290–18 304, 2024.
- [9]. C. Deina et al., “Decision analysis framework for predicting no-shows to appointments using machine learning algorithms,” *BMC Health Services Research*, vol. 24, no. 1, pp. 1–12, 2024.
- [10]. S. Khalilpourazari et al., “Designing an efficient blood supply chain network in crisis: neural learning, optimization and case study,” *Annals of Operations Research*, vol. 283, pp. 335–363, 2019.
- [11]. T. Niu et al., “A review of optimization studies for system appointment scheduling,” *Axioms*, vol. 13, no. 1, p. 52, 2024.
- [12]. A. Ala and F. Chen, “Appointment scheduling problem in complexity systems of the healthcare services: A comprehensive review,” *Journal of Healthcare Engineering*, vol. 2022, 2022.
- [13]. R. Vassalli, J. Vojnovic’, and G. Magnini, “Modeling blood donation behavior and social cohesion in europe,” *SSRN*, 2025, preprint.
- [14]. M. A. Mohammed, “Use of machine learning in optimizing medical appointment schedules,” *International Journal of Computer Science and Information Technologies*, vol. 14, no. 1, 2023.
- [15]. A. M. Alturaiki et al., “A smart chatbot for interactive management in beta thalassemia patients,” *International Journal of Telemedicine and Applications*, vol. 2022, 2022.
- [16]. R. D. Ismael, H. A. Hussein, and M. M. Salih, “A systematic review on smart blood bank system: Taxonomy, motivations, challenges, study di- rections and recommendations,” *Journal of Algebraic Statistics*, vol. 13, no. 3, pp. 2490–2510, 2022.
- [17]. S. AlZu’bi, D. Aqel, and A. Mughaid, “Recent intelligent approaches for managing and optimizing smart blood donation process,” in *2021 International Conference on Information Technology (ICIT)*. IEEE, 2021, pp. 386–391.
- [18]. J. C. Matulis and R. G. McCoy, “Patient-centered appointment schedul- ing: a call for autonomy, continuity, and creativity,” *Journal of General Internal Medicine*, vol. 35, no. 9, pp. 2743–2746, 2020.
- [19]. L. Z. Song, Y. and N. Sahoo, “Matching returning donors to projects on philanthropic crowdfunding platforms,” *SSRN Preprint*, N/A, year not available in source.