Cordoba International Journal of Academic Research (CIJAR) Online, open access

Volume 1, Issue 4 (October – December 2026), pp. 133–149 | Published: 15 October 2026
Publisher: Wexgrand Academic Press (a division of Wexgrand Private Limited)
Article URL: https://cijar.org/article/explainable-ai-for-decision-support-in-blood-bank-operations
Cite as: Linga Abidha Linga Vasagam (2026). Explainable AI for Decision Support in Blood Bank Operations. *Cordoba International Journal of Academic Research*, 1(4), 133–149.

Explainable AI for Decision Support in Blood Bank Operations

Linga Abidha Linga Vasagam
B.Sc, Data Science
SIES College Of Arts, Science & Commerce (Autonomous), India

Abstract

Artificial Intelligence (AI) and Machine Learning (ML) are increasingly being applied in blood bank operations to support demand forecasting, inventory management, shortage prediction, and resource allocation. However, the complexity of many AI models creates challenges in understanding and trusting their predictions. Explainable Artificial Intelligence (XAI) can address this issue by providing understandable explanations for AI-driven recommendations. This paper presents a systematic review of AI and ML applications in blood bank decision support, with particular focus on explainability, transparency, and human-centered decision-making. The review examines existing approaches, their benefits and limitations, and identifies research gaps in the development of explainable decision-support systems for blood bank operations. Based on the reviewed literature, a conceptual framework integrating operational data, AI/ML models, an explainability layer, decision-support outputs, and human review is proposed. The framework emphasizes transparency, validation, and human oversight to support responsible adoption of AI in blood bank management.

Key Words: *Artificial Intelligence (AI), Machine Learning (ML), Explainable AI (XAI), Blood Bank Management, Decision Support Systems, Blood Inventory Management, Healthcare AI, Systematic Review*

I. INTRODUCTION

Blood banks operate in an environment where supply and demand can change quickly. Decisions concerning blood-component availability, inventory levels, expiry, shortage response, and allocation must be made using information that may arrive from multiple operational sources. The

consequences of poor decisions can extend beyond routine inefficiency because blood products are time-sensitive resources and some clinical requirements cannot be postponed.

Artificial intelligence has increasingly been studied as a means of supporting healthcare decisions. Machine-learning systems can identify patterns in historical data and generate forecasts, classifications, or risk estimates. In medicine, however, the usefulness of a model is not determined only by predictive performance. The people using the result also need to understand the circumstances under which the model produced it, especially when the output affects a high-stakes workflow. [1]

This concern motivates the use of Explainable Artificial Intelligence. XAI includes techniques that help describe the relationship between input information and model output. Methods such as LIME and SHAP can provide local explanations for individual predictions, while other approaches examine model behaviour more broadly. [2] The purpose of an explanation in a blood-bank setting is not to replace professional judgement. Rather, it is to provide additional evidence that allows staff to inspect an AI-supported recommendation and decide whether it is appropriate for the current operational context.

The research problem therefore lies at the intersection of three areas: operational decision-making in blood banks, AI-based prediction, and human understanding of model output. Research on healthcare AI and XAI provides useful methodological foundations, but these ideas need to be considered in relation to the particular characteristics of blood-bank operations, including time sensitivity, inventory constraints, data quality, and human accountability.

A. Problem Statement

AI-based decision support can produce useful forecasts or alerts while remaining difficult for operational users to interpret. A system may indicate that a shortage is likely without showing which variables contributed to that conclusion. Similarly, an inventory recommendation may be mathematically reasonable but difficult for staff to assess if the system does not expose the relevant evidence. This creates a practical gap between model output and human decision-making.

B. Purpose of the Study

The purpose of this study is to examine the role of Explainable AI in decision-support applications related to blood-bank operations. The study focuses on how explanations can make AI-supported outputs more transparent, inspectable, and usable while recognising that explanations do not

automatically guarantee model correctness or clinical safety.

C. Research Objectives

The study aims to:

1. identify AI and ML applications relevant to blood-bank operational decisions;
2. examine the types of decisions supported by these systems;
3. identify XAI techniques that can explain model predictions and recommendations;
4. examine reported benefits and limitations of explainability in healthcare decision support; and
5. identify research gaps that can guide future blood-bank decision-support systems.

D. Research Questions

The review is organised around the following questions:

1. What AI/ML applications are used to support decisions in blood-bank or transfusion-related operations?
2. What types of operational decisions are supported by these applications?
3. Which XAI techniques are used to make AI-supported outputs understandable?
4. What benefits and challenges of XAI are reported for healthcare decision support?
5. What gaps remain for explainable AI in blood-bank operational decision support?

II. BACKGROUND AND CONTEXT

A. AI in Healthcare Decision Support

AI-based systems are increasingly used to process complex healthcare information, identify patterns, assist prediction, and support decisions. Rajkomar et al. describe machine learning as a set of approaches that can learn from healthcare data and support tasks that would otherwise require substantial manual analysis. [3] Topol similarly argues that the most useful role for AI in medicine is not simply automation but an interaction between computational systems and human expertise. [4]

For blood-bank operations, the relevant data may include historical

demand, current stock, blood group, component type, issue and return records, expiry dates, collection information, and emergency requests. Such data can support forecasting and operational monitoring. However, the availability of data does not by itself establish that a model is suitable for deployment. Data quality, representativeness, missing values, and changes in operating conditions can affect model behaviour.

B. Blood Bank Operations

Blood banks manage a set of processes that connect collection, testing, processing, storage, inventory control, distribution, and transfusion support. Unlike many ordinary inventory settings, blood components have biological and regulatory constraints. Some products have limited shelf lives, demand can vary by blood group and clinical situation, and emergency requirements can change the normal allocation pattern.

These characteristics create several possible decision-support tasks. Forecasting may be used to estimate future demand; inventory models may help identify potential shortages or excess stock; monitoring systems may flag products approaching expiry; and allocation tools may assist staff in evaluating available units against current requirements. In each case, an AI system would be supporting a human workflow rather than making an independent clinical decision.

C. Explainable Artificial Intelligence

XAI is concerned with making the behaviour or output of an AI system understandable to people. Arrieta et al. describe XAI as a broad area containing different definitions, taxonomies, and explanation methods rather than a single algorithm. [5] Doshi-Velez and Kim emphasise the need for a rigorous science of interpretability and caution against treating interpretability as a single property. [6]

Local explanation methods are particularly relevant to decision support because they can describe why a particular prediction was made. LIME approximates a complex model locally to explain an individual prediction. [7] SHAP assigns contribution values to features based on Shapley-value concepts and can be used to examine feature contributions for individual predictions and groups of predictions. [8] These methods can be useful for operational review, but their explanations should not be confused with proof that a model is correct.

D. Why Explainability Matters in Blood Banks

A blood-bank employee may need to answer practical questions such as: Why has a shortage alert appeared? Which variables contributed to the

forecast? Why is a particular component being prioritised? What information is the recommendation based on? An explanation layer can help turn a model output into information that can be inspected alongside operational evidence.

Explainability can therefore support transparency and review, but it also introduces design questions. An explanation that is technically accurate may be too complex for a busy operational environment. Conversely, an overly simplified explanation may hide important uncertainty. The useful explanation is therefore not necessarily the most detailed one; it is the one that provides relevant information in a form that supports appropriate human judgement.

III. RELATED WORK

Research on XAI has developed alongside broader work on machine learning in healthcare. The literature commonly distinguishes between model-specific and model-agnostic approaches, local and global explanations, and inherently interpretable models versus post-hoc explanations. [9]

Ribeiro et al. introduced LIME as a model-agnostic method for explaining individual predictions by fitting an interpretable approximation around a particular instance. [10] Lundberg and Lee proposed SHAP as a unified framework for assigning importance values to features in a prediction. [11] These methods are relevant to decision-support research because they can expose the contribution of input variables without requiring users to inspect the internal mathematical structure of a complex model.

Healthcare applications introduce additional requirements. Amann et al. argue that explainability in healthcare needs to be considered together with factors such as transparency, accountability, safety, and the intended users of an AI system. [12] Rudin also argues that in high-stakes settings, researchers should consider interpretable models directly where they can achieve the required performance, rather than assuming that a black-box model followed by a post-hoc explanation is always the appropriate solution. [13]

These observations are relevant to blood-bank operations. If the operational task can be handled by a transparent statistical or rule-based model without sacrificing useful performance, such a model may be easier to validate. Where more complex machine-learning models are necessary, an explanation layer may provide additional insight. The choice should therefore depend on the task, data, risk, users, and validation evidence rather than on the use of XAI as a label alone.

Table 1: Examples of AI-supported blood-bank decision areas

Decision area	Possible AI output	Role of explanation
Demand planning	Forecast of future component demand	Show influential historical patterns, seasonality, and relevant inputs.
Inventory monitoring	Shortage or excess-stock risk	Identify factors contributing to the alert and provide supporting context.
Expiry management	Risk of units approaching expiry	Indicate the variables or patterns associated with the risk estimate.
Resource allocation	Suggested prioritisation or matching	Provide reasons and constraints considered by the model.
Emergency support	Rapid risk or availability estimate	Present the evidence and uncertainty needed for human review.

IV. METHODOLOGY

A. Research Design

This study adopts a systematic literature review design. The review is intended to identify, organise, and synthesise research concerning AI, XAI, and decision support in blood-bank and closely related transfusion settings. PRISMA 2020 is used as the reporting framework because it provides a structured approach for documenting identification, screening, eligibility, and inclusion of studies. [14]

B. Search Strategy

The search strategy combines three concept groups: explainable or interpretable AI; blood-bank or transfusion contexts; and decision-support or operational tasks. A draft search expression is:

(*"explainable AI"* OR *"explainable artificial intelligence"* OR XAI OR *"interpretable machine learning"*) AND

(*"blood bank"* OR *"blood banking"* OR *"blood inventory"* OR *"blood transfusion"* OR *"transfusion medicine"*) AND

(*"decision support"* OR *"decision making"* OR *forecasting* OR *inventory* OR *shortage* OR *wastage* OR *allocation*)

The intended databases are PubMed, IEEE Xplore, Scopus, ScienceDirect, SpringerLink, and Google Scholar as a supplementary discovery source. Search terms may be adapted to the syntax supported by each database while preserving the three main concepts.

C. Eligibility Criteria

Studies will be considered when they address AI/ML in blood-bank, transfusion, or closely related operational healthcare contexts and provide information relevant to decision support or explainability. Studies focused only on unrelated clinical diagnosis, purely theoretical XAI without an operational

healthcare connection, editorials without substantive evidence, and duplicate records will be excluded.

Table 2: Planned eligibility criteria

Include	Exclude
AI/ML studies addressing blood-bank, blood-supply, or transfusion-related decisions	Studies with no relevant blood or transfusion context.
Studies reporting a decision-support task, prediction, recommendation, classification, or operational analysis	Papers unrelated to an operational or decision-support problem.
Studies that discuss interpretability, explanation, transparency, or an XAI technique	General AI commentary without a substantive research contribution.
Peer-reviewed research and relevant scholarly conference papers with sufficient methodological detail	Duplicates, inaccessible records, or papers lacking enough information for extraction.

D. Study Selection and Data Extraction

Records will be screened first by title and abstract and then by full text. The final set will be recorded using a PRISMA flow, including reasons for exclusion at the full-text stage. For each included study, the extraction form will capture the research objective, operational application, dataset, AI/ML method, explainability approach, decision supported, main findings, limitations, and research gap.

E. Quality Assessment

Quality assessment will consider whether a study clearly describes its research objective, data source, modelling method, evaluation procedure, explanation method, and limitations. The assessment is intended to distinguish evidence that can support a conclusion from work that is primarily conceptual. No numerical quality score is reported in this manuscript until the actual included studies have been screened and assessed.

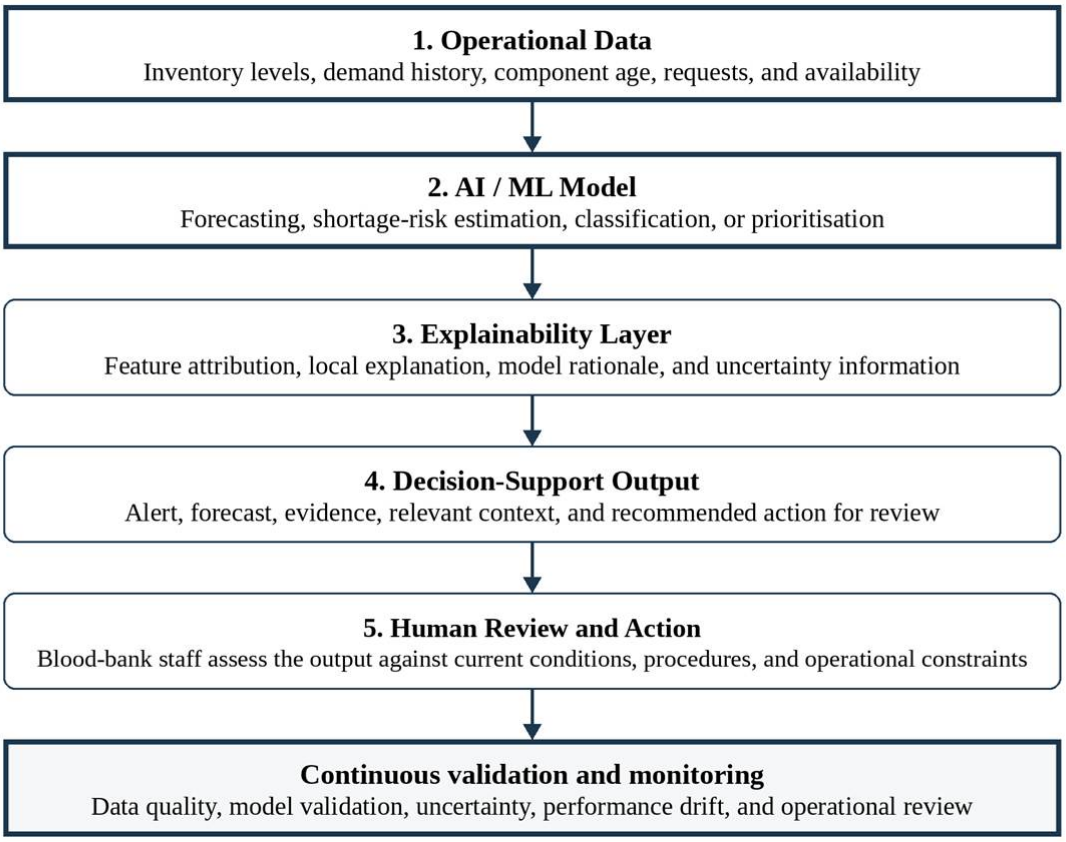
F. Synthesis Approach

The evidence will be synthesised thematically rather than by combining incompatible performance measures. Themes will include operational problem, model type, explanation method, user role, benefits, limitations, validation, and implementation requirements. This approach is appropriate because the expected literature may contain different datasets, prediction targets, metrics, and operational settings.

V. PROPOSED EXPLAINABLE DECISION-SUPPORT FRAMEWORK

The conceptual framework translates the research questions into a practical information flow. It is a research model rather than a claim that a validated production system has already been implemented.

Figure 1: Conceptual framework for explainable AI-based decision support in blood-bank operations. The framework separates model prediction from human decision-making and treats validation and monitoring as continuous activities.



The first stage contains operational information such as current inventory, historical demand, component characteristics, and requests. The second stage produces a forecast, risk estimate, classification, or other model output. The XAI layer then provides an explanation appropriate to the selected model. Finally, the output is presented as decision support to a human user. Validation and monitoring are treated as continuous activities rather than as one-time steps.

The framework deliberately separates prediction from decision. An AI system can estimate the probability of a shortage, but the operational response may also depend on current stock, incoming supply, clinical urgency, and local procedures. Therefore, the final action should remain subject to human review and organisational rules.

A. Explanation Design

An explanation interface could contain three elements: the model output, the most influential factors, and relevant supporting context. For example, a shortage-risk alert might show the predicted risk together with recent demand, current inventory, and the variables that contributed most strongly to the estimate. Where the model permits it, uncertainty or confidence information should also be visible.

B. Human Oversight

The framework treats explainability as a support mechanism for review rather than as a substitute for expertise. Users should be able to inspect an alert, compare it with operational evidence, and reject or investigate a recommendation when the circumstances do not match the model assumptions. This principle is consistent with the broader healthcare literature emphasising human responsibility and appropriate use of AI outputs. [15]

VI. KEY FINDINGS AND LITERATURE-BASED INSIGHTS

The literature establishes several points that are relevant when XAI is considered for blood-bank decision support. First, predictive performance alone does not describe whether an AI system is suitable for practical healthcare use. Model behaviour, data quality, user interaction, and accountability also matter. [16]

Second, explainability is not a single technique. LIME and SHAP provide different approaches to explaining individual predictions, while intrinsically interpretable models may provide a different form of transparency. [17] The appropriate method therefore depends on the model, the decision being supported, and the information the user needs.

Third, explanations should be designed for the intended user. A data scientist may understand feature-attribution values that are not meaningful to a blood-bank operator. A useful system therefore needs to connect technical explanations to operational concepts such as stock level, recent demand, expiry, or request urgency.

Fourth, explanation quality should not be treated as evidence of model correctness. A model can produce a plausible-looking explanation for an incorrect prediction. XAI should therefore be accompanied by validation, monitoring, and data-quality checks. This distinction is particularly important in high-stakes environments, where an understandable model can still be unsafe if its underlying predictions are unreliable.

Finally, the blood-bank setting creates an opportunity for decision support that is more transparent than fully automated decision-making. Forecasts and risk alerts can be used to draw attention to situations requiring review while allowing staff to combine model output with operational knowledge.

A. Operational Interpretation

The practical value of XAI can be viewed through four questions: What did the model predict? Why did it produce that output? What information was available to the model? What should the user verify before acting? Organising an interface around these questions can make an explanation more useful than displaying a technical score alone.

B. Reliability and Validation

Reliability requires attention to both the model and the data pipeline. Changes in demand patterns, recording practices, stock policies, or external events can alter the relationship learned from historical data. A model should therefore be monitored after deployment, and explanations should be reviewed when the data distribution changes.

VII. DISCUSSION

The central argument emerging from the literature is that XAI should be considered as part of a broader decision-support design rather than added as a visual feature after model development. In blood-bank operations, the user needs to understand both the computational result and the operational context in which it is being used.

A forecasting model may identify a likely increase in demand, but the explanation should help the user determine whether the increase is associated with recent trends, particular blood groups, seasonal effects, or another measurable pattern. Similarly, an expiry-risk system should make it possible to understand why a product has been prioritised for attention. These examples illustrate the difference between explaining a model and supporting a decision.

The literature also suggests that there is a trade-off between complexity and interpretability. Highly flexible models can capture nonlinear relationships, but their outputs may require post-hoc explanation. Simpler models may be easier to inspect directly. Rudin’s argument for interpretable models in high-stakes settings is therefore relevant when selecting an approach for blood-bank applications. [18]

From a system-design perspective, explainability should be connected to validation. A useful architecture should record the model version, input data used for a prediction, explanation output, and subsequent human action where

appropriate. Such records can support auditing and help identify cases in which the system's recommendation disagreed with operational judgement.

A. Implications for Blood Bank Decision Support

The findings indicate that future systems should prioritise clear operational questions rather than starting with a particular AI algorithm. The sequence should be: define the decision, identify the available data, select a model appropriate to the task, evaluate its performance, choose an explanation method, and test whether users can interpret the explanation correctly. This approach keeps the technology subordinate to the operational problem.

VIII. RESEARCH GAPS

Several gaps remain at the intersection of XAI and blood-bank operations. One gap is the limited integration of explanation design with operational workflows. Many XAI studies evaluate explanations as technical outputs, whereas a blood-bank user needs explanations that are connected to actual inventory and supply decisions.

A second gap concerns evaluation with real users. An explanation method may appear informative to researchers but still fail to help operational staff make better decisions. Future work should therefore include usability studies and human-factors evaluation alongside conventional model metrics.

A third gap is the treatment of uncertainty. A feature-attribution explanation can show why a model generated an output without showing how stable that output is under changing conditions. Blood-bank decision support should therefore investigate uncertainty communication, data drift, and model monitoring together with explanation.

A fourth gap is evidence from realistic datasets. Blood-bank datasets can contain privacy restrictions, institutional differences, missing values, and changing operating policies. Reproducible research will require careful documentation of these limitations and transparent evaluation procedures.

IX. PROPOSED EVALUATION APPROACH

A future implementation should be evaluated at more than one level. Model performance is necessary, but it is not sufficient to establish whether an explainable decision-support system is useful in a blood-bank workflow. The evaluation should therefore combine technical, explanation, and human-centred measures.

At the model level, appropriate forecasting or classification measures should be selected according to the operational task. For a demand-forecasting

problem, measures such as MAE or RMSE may be appropriate; for a shortage-risk classifier, precision, recall, F1-score, and calibration may be more informative. The selected metric should reflect the consequences of false alerts and missed alerts rather than being chosen only for convenience.

At the explanation level, future work should examine whether the explanation is stable, understandable, and consistent with the model behaviour. A useful explanation should help a user identify the major factors behind an output without implying a causal relationship that the model has not established. Where different explanation methods are compared, their agreement and stability should also be examined.

At the user level, a controlled evaluation could compare decisions made with and without explanation support. Measures could include task completion time, identification of incorrect alerts, confidence in the decision, and the user's ability to correctly describe why the model produced an output. This would provide stronger evidence of practical usefulness than presenting explanation visualisations alone.

Table 3: Suggested evaluation dimensions for a future prototype

Level	Example measures	Purpose
Model	MAE, RMSE, precision, recall, F1, calibration	Determine whether the prediction or risk estimate is sufficiently reliable for the intended task.
Explanation	Stability, consistency, feature relevance, user comprehension	Determine whether the explanation accurately and usefully communicates model behaviour.
Workflow	Decision time, missed alerts, unnecessary alerts, task accuracy	Determine whether the system supports the operational process rather than adding avoidable burden.
Human factors	Trust calibration, confidence, explanation recall, usability	Examine whether users understand and appropriately act on the AI-supported information.

The evaluation should also include comparison with a reasonable baseline, such as an existing forecasting method or a transparent statistical model. This would allow the study to determine whether the additional complexity of an AI/XAI system provides a meaningful operational benefit.

X. ETHICAL AND DATA GOVERNANCE CONSIDERATIONS

Blood-bank information may contain sensitive operational and health-related data. Future implementation should therefore apply data minimisation, access control, de-identification where appropriate, secure storage, and clear governance procedures. The research use of data should follow the

requirements of the institution and applicable ethical review processes.

An explainable system also creates an auditability opportunity. A future platform could retain the model version, relevant input snapshot, explanation, and final human action for each important recommendation, subject to privacy and retention requirements. Such records could help investigate unusual decisions and support model monitoring.

Another consideration is automation bias. Users may give excessive weight to a system recommendation simply because it is presented as an AI output. The interface should therefore communicate uncertainty and encourage appropriate verification. An explanation should support critical review, not make the recommendation appear more authoritative than the underlying evidence warrants.

The system should also distinguish between correlation and operational causation. If a feature contributes strongly to a prediction, that does not mean changing the feature would necessarily change the outcome. Presenting feature attribution as a causal explanation could lead to incorrect operational decisions. Clear terminology and user training are therefore part of responsible deployment.

XI. LIMITATIONS

This paper is a literature-based study and does not present a newly trained blood-bank prediction model or claim measured improvement in operational outcomes. The proposed framework is conceptual. The eventual systematic review results will depend on the databases searched, search terms, eligibility decisions, and availability of relevant studies.

A further limitation is that XAI methods do not have a universally accepted definition of explanation quality. Different methods may produce different explanations for the same prediction, and users may interpret those explanations differently. These limitations mean that explainability should be evaluated together with model validity, user needs, and the intended decision context.

XII. FUTURE RESEARCH

Future research can develop and evaluate a prototype decision-support platform using real or appropriately de-identified blood-bank data. The system could combine inventory monitoring, demand forecasting, shortage alerts, and expiry-risk monitoring with an explanation interface.

A second direction is comparative evaluation of interpretable models and

post-hoc XAI methods. Rather than assuming that one method is superior, future studies could compare transparency, predictive performance, explanation stability, and user understanding for the same operational task.

User-centred evaluation is also important. Blood-bank staff could be involved in testing whether explanations are understandable, whether they help identify data problems, and whether they support appropriate review of AI recommendations. Such evaluation would provide evidence about practical usefulness rather than only technical explainability.

Finally, future systems should examine monitoring after deployment. Data drift, changes in demand, and changes in operational procedures can affect model reliability. Continuous monitoring combined with explanation logging could provide a basis for safer long-term use.

XIII. CONCLUSION

Explainable AI has the potential to strengthen AI-based decision support in blood-bank operations by making model outputs easier to inspect and discuss. The literature shows that XAI is broader than a single explanation algorithm and that the choice between interpretable models and post-hoc explanations depends on the task, model, user, and risk context.

For blood-bank operations, the most useful role of XAI is likely to be as a bridge between prediction and human judgement. Forecasts, shortage alerts, and other model outputs can provide useful evidence, while explanations can show the factors behind those outputs and help users decide what should be checked before action is taken. The proposed framework therefore places operational data, AI/ML prediction, explanation, validation, and human oversight in one workflow.

The study does not claim that explainability by itself makes an AI system reliable. Instead, it identifies explainability as one component of a broader approach that includes data quality, model validation, uncertainty, monitoring, and responsible human use. These considerations provide a foundation for future empirical work on explainable decision-support systems for blood-bank operations.

ACKNOWLEDGMENT

The author expresses sincere gratitude to the Department of Data Science, SIES College of Arts, Science and Commerce (Empowered Autonomous), Sion (W), Mumbai, for providing the academic environment and support required for this research work. The author also thanks the project guide, Assistant Professor Smita Mangesh Junnarkar, for guidance and

encouragement throughout the study.

Endnotes

[1] . A. Rajkomar, J. Dean, and I. Kohane, “Machine learning in medicine,” *New England Journal of Medicine*, vol. 380, pp. 1347–1358, 2019, doi: 10.1056/NEJMra1814259.; E. J. Topol, “High-performance medicine: The convergence of human and artificial intelligence,” *Nature Medicine*, vol. 25, pp. 44–56, 2019, doi: 10.1038/s41591-018-0300-7.

[2] . M. T. Ribeiro, S. Singh, and C. Guestrin, “Why should I trust you? Explaining the predictions of any classifier,” in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016, pp. 1135–1144, doi: 10.1145/2939672.2939778.; S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in *Advances in Neural Information Processing Systems*, vol. 30, 2017.; A. B. Arrieta et al., “Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI,” *Information Fusion*, vol. 58, pp. 82–115, 2020, doi: 10.1016/j.inffus.2019.12.012.

[3] . A. Rajkomar, J. Dean, and I. Kohane, “Machine learning in medicine,” *New England Journal of Medicine*, vol. 380, pp. 1347–1358, 2019, doi: 10.1056/NEJMra1814259.

[4] . E. J. Topol, “High-performance medicine: The convergence of human and artificial intelligence,” *Nature Medicine*, vol. 25, pp. 44–56, 2019, doi: 10.1038/s41591-018-0300-7.

[5] . A. B. Arrieta et al., “Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI,” *Information Fusion*, vol. 58, pp. 82–115, 2020, doi: 10.1016/j.inffus.2019.12.012.

[6] . F. Doshi-Velez and B. Kim, “Towards a rigorous science of interpretable machine learning,” arXiv:1702.08608, 2017.

[7] . M. T. Ribeiro, S. Singh, and C. Guestrin, “Why should I trust you? Explaining the predictions of any classifier,” in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016, pp. 1135–1144, doi: 10.1145/2939672.2939778.

[8] . S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[9] . A. B. Arrieta et al., “Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI,” *Information Fusion*, vol. 58, pp. 82–115, 2020, doi: 10.1016/j.inffus.2019.12.012.; C. Rudin, “Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead,” *Nature Machine Intelligence*, vol. 1, pp. 206–215, 2019, doi: 10.1038/s42256-019-0048-x.

[10] . M. T. Ribeiro, S. Singh, and C. Guestrin, “Why should I trust you? Explaining the predictions of any classifier,” in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016, pp. 1135–1144, doi: 10.1145/2939672.2939778.

[11] . S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[12] . J. Amann, A. Blasimme, H. Vayena, D. Frey, and V. I. Madai, “Explainability for artificial intelligence in healthcare: A multidisciplinary perspective,” *BMC Medical Informatics and Decision Making*, vol. 20, 2020, doi: 10.1186/s12911-020-01332-6.

[13] . C. Rudin, “Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead,” *Nature Machine Intelligence*, vol. 1, pp. 206–215, 2019, doi: 10.1038/s42256-019-0048-x.

[14] . M. J. Page et al., “The PRISMA 2020 statement: An updated guideline for reporting systematic reviews,” *BMJ*, vol. 372, n71, 2021, doi: 10.1136/bmj.n71.

[15] . J. Amann, A. Blasimme, H. Vayena, D. Frey, and V. I. Madai, “Explainability for artificial intelligence in healthcare: A multidisciplinary perspective,” *BMC Medical Informatics and Decision Making*, vol. 20, 2020, doi: 10.1186/s12911-020-01332-6.; E. J. Topol, “High-performance medicine: The convergence of human and artificial intelligence,” *Nature Medicine*, vol. 25, pp. 44–56, 2019, doi: 10.1038/s41591-018-0300-7.

[16] . J. Amann, A. Blasimme, H. Vayena, D. Frey, and V. I. Madai, “Explainability for artificial intelligence in healthcare: A multidisciplinary perspective,” *BMC Medical Informatics and Decision Making*, vol. 20, 2020, doi: 10.1186/s12911-020-01332-6.; E. J. Topol, “High-performance medicine: The convergence of human and artificial intelligence,” *Nature Medicine*, vol. 25, pp. 44–56, 2019, doi: 10.1038/s41591-018-0300-7.

[17] . M. T. Ribeiro, S. Singh, and C. Guestrin, “Why should I trust you? Explaining the predictions of any classifier,” in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and*

Data Mining, 2016, pp. 1135–1144, doi: 10.1145/2939672.2939778.; S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.; C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature Machine Intelligence*, vol. 1, pp. 206–215, 2019, doi: 10.1038/s42256-019-0048-x.

[18] . C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," *Nature Machine Intelligence*, vol. 1, pp. 206–215, 2019, doi: 10.1038/s42256-019-0048-x.

© 2026 The Author(s). Published in *Cordoba International Journal of Academic Research*.
This work is licensed under a **Creative Commons Attribution 4.0 International License (CC BY 4.0)**.
Published by Wexgrand Academic Press (a division of Wexgrand Private Limited), No. 4/461, Suite No. 1331, Valamkottil Towers, Thrikkakara, Ernakulam, Kerala – 682021, India.