International Journal on Advanced Science Engineering Information Technology

Vol.15 (2025) No. 1
ISSN: 2088-5334

Generative AI-Powered Synthetic Data for Enhancing Predictive Analytics in Blood Donation Supply Management: A Comparative Study of Machine Learning Models

Koh Chee Hong [a], Thong Chee Ling [a], Shayla Islam [a], Raenu AL Kolandaisamy [a], Abdul Samad Bin Shibghatullah [b], Nazirul Nazrin Bin Shahrol Nidzam [c,*], Samer Sarsam [d], Halimah Badioze Zaman [e]

[a] *Institute of Computer Science & Digital Innovation (ICSDI), UCSI University, Kuala Lumpur, Malaysia*
[b] *College of Computing & Informatics, Universiti Tenaga Nasional, Putrajaya Campus, Selangor, Malaysia*
[c] *Institute of Informatics and Computing in Energy, Universiti Tenaga Nasional, Putrajaya Campus, Selangor, Malaysia*
[d] *Faculty of Business & Law, Coventry University, Priory St, Coventry, United Kingdom*
[e] *Tan Sri Leo Moggie Distinguished Chair in Energy Informatics, IICE, Universiti Tenaga Nasional (UNITEN), Malaysia*

Corresponding author: [*]nazirul.nazrin@uniten.edu.my

***Abstract*— Maintaining a sufficient and timely blood supply is an urgent and critical challenge in public health, where even minor miscalculations can lead to life-threatening shortages. This study evaluates the performance of machine learning models to improve blood donation forecasting. Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) generated synthetic datasets that mirror real-world donation patterns to address data scarcity and variability issues. Leveraging transactional data from the Blood Bank Information System (BBISv2), a blood tracking system used by 22 main blood collection sites under the Ministry of Health (MoH) in Malaysia, 50 synthetic datasets were created and validated to ensure consistency with real data. The synthetic data showed minimal deviations from real data across key metrics, including mean (differences under 10%), variance (1 to 2 units), and skewness and kurtosis (0.03 or less). Among the models, the Random Forest algorithm demonstrated the highest performance, achieving an accuracy of 98.7%, a precision of 0.91, and an Area Under the Receiver Operating Characteristic (AUC-ROC) score of 0.92, making it the most reliable for predicting blood donation rates. Linear Regression also performed well, with an accuracy of 98.6%, while Neural Networks and Support Vector Machines (SVM) showed lower performance. This research provides a valuable tool for optimizing blood donation strategies, particularly in scenarios where real data is limited. Integrating validated synthetic data offers a novel approach for enhancing resource management in healthcare, ensuring reliable blood supply during high-demand periods.**

***Keywords*—Blood donation forecasting; predictive analytics and visualization; generative AI; random forest; public health.**

Manuscript received 7 Jan. 2024; revised 11 Sep. 2024; accepted 15 Nov. 2024. Date of publication 28 Feb. 2025.

I. INTRODUCTION

Blood shortages are a critical challenge in healthcare systems, particularly in emergency care and chronic illness treatments, as noted by Sethuraman et al. [1]. Predicting blood donation rates is complex, with numerous dynamic factors leading to inefficiencies and potential shortages, as discussed by Maheshwari et al. [2]. Traditional forecasting methods often fail to capture these variables, increasing risks for patients needing transfusions, as highlighted by VanVactor [3]. Clauson et al. [4] stressed the importance of predictive analytics, and Scala and Lindsay [5] argued for more robust models during crises. Various studies have explored factors influencing donation rates. Lin et al. [6] examined meteorological impacts, while Intharanut et al. [7] studied donor characteristics. Ahamed et al. [8] used regression models for donor forecasting, and Shih and Rajendran [9] compared time series with machine learning algorithms. However, these models often miss complex, non-linear factors and lack adaptability, as Thijsen et al. [10] and Adhikari et al. [11] noted.

This study introduces a hybrid model that integrates Generative AI (GANs and VAEs) with traditional machine learning methods like the PCC and Random Forest. Kumar et al. [12] suggest that this integration enhances prediction accuracy by generating synthetic data that reflects real-world

conditions. The model captures complex interactions between CSFs, providing a more robust solution for blood supply forecasting, as explored by Alhamda and Rahman [13], Yu et al. [14], and Chansamut [15]. Using data from Malaysia's BBISv2 system validates the model in real-world settings. Validating synthetic data with GANs and VAEs ensures it mirrors real patterns accurately, as proposed by Reddy [16] and Krishnan and Khastgir [17]. This study contributes to healthcare analytics, enhancing blood supply management and potentially saving lives, as discussed by Longaray et al. [18], anchoring on the following research objectives:

a. Develop a hybrid predictive model using Generative AI and traditional machine learning for accurate blood supply forecasting.
b. Validate synthetic data generated by GANs and VAEs to align with real-world patterns.
c. Compare the hybrid model against traditional models to show its superior performance.
d. Offer a tool for improved blood supply management in healthcare.

By achieving these objectives, the research bolsters blood supply chain resilience, aligning with public health improvement goals, as emphasized by Haw et al. [19] and Pahune and Rewatkar [20]. Accurate blood donation forecasting is essential to prevent shortages, as Bou Assi et al. [21] highlighted. Donation rates are influenced by complex factors, including socioeconomic conditions and donor behavior, as noted by Sethuraman et al. [1]. Despite insights from Lin et al. [6] and Intharanut et al. [7], models still struggle with accuracy under non-standard conditions, indicating gaps in the literature, as pointed out by VanVactor [3]. Traditional models, such as those discussed by Shih and Rajendran [9] and Thakur et al. [22], fail to capture complex interactions in blood donation patterns. They overlook perishability and dynamic supply-demand factors, limiting predictive power, as Ahamed et al. [8] and Krishnan and Khastgir [17] noted. These gaps necessitate advanced models, as highlighted by Chansamut [15] and Kumar et al. [12].

CSFs play a crucial role in blood donation supply chains. As Kumar et al. [12] and Wood [23] stated, state population impacts the donor pool. Donation rates vary with the day of the week, influenced by cultural norms, as Thijsen et al. [10] and Lin et al. [6] discussed. Public events and holidays can affect donation rates, as shown by Thakur et al. [22] and Bou Assi et al. [21]. Historical trends and blood group distribution are also essential for forecasting, as noted by Ahamed et al. [8], Maheshwari et al. [2], and Gupta et al. [24].

Past research established a foundation but lacked advanced machine learning integration, as Scriffignano [25] stated. Models like Lin et al. [6] and Thakur et al. [22] rely on historical data, missing complex patterns. They fail to address data scarcity, a limitation noted by Maheshwari et al. [2]. Advanced models generating synthetic data are necessary, as argued by Afrashtehfar and Abu-Fanas [26]. Technologies like blockchain, discussed by Clauson et al. [4], and AI in decision-making, outlined by Adhikari et al. [11], offer additional solutions. As Nsikan et al. [27] emphasized, real-time adaptability is crucial for managing blood donation patterns. Generative AI techniques like GANs and VAEs can address data scarcity, as Labib and Gharib [28] discussed. Despite their potential, their application in blood donation forecasting remains underexplored, as noted by Reddy [16]. Patel and Malik [29] highlighted how Generative AI could revolutionize predictive analytics. Hashizume et al. [30] stressed AI's role in comprehensive data integration, and Al-Ma'aitah [31] emphasized the adaptability of AI-driven models in healthcare. Integrating Generative AI with machine learning, like PCC and Random Forest, improves forecasting accuracy, as argued by Hashizume et al. [30]. This approach overcomes dataset limitations, aligning with Pahune and Rewatkar [20] and Scala and Lindsay [5]. Selvaraj [32] emphasized Generative AI's power in data integration, while Adhikari et al. [11] and Malik and Naudiyal [33] noted the importance of robust, scalable AI models.

Hybrid models combining GANs, VAEs, and predictive machine learning models offer a solid framework for blood donation prediction, as Lemon [34] suggested. PCC identifies influential CSFs, enhancing model interpretability, as Krishnan and Khastgir [17] discussed. Random Forest reduces overfitting, as noted by Chai et al. [35]. Combining AI techniques improves predictive accuracy and resilience [36], as Nsikan et al. [27] and Al-Ma'aitah [31] emphasized.

Validating synthetic data and evaluating models is crucial, as Essila [37] stated. Metrics like mean, variance, and the Kolmogorov-Smirnov test assess data alignment, as Adhikari et al. [11] noted. The AUC-ROC score is vital for performance evaluation, balancing sensitivity and specificity, as Reddy [16] and Lee et al. [38] discussed. Rigorous validation in predictive analytics is essential, ensuring adaptability, as Dhan and Kumar [39] underscored. Hybrid models integrating machine learning with Generative AI represent a significant advancement in blood donation forecasting, as highlighted by Selvaraj [32]. They address traditional models' limitations, offering precise insights for resource allocation, as Gupta et al. [24] argued. Venugopal and Venugopal [40] validated the effectiveness of ensemble learning in healthcare, ensuring efficient resource management and adequate blood supply.

II. Materials and Method

A. Research Design

This study developed a predictive model for blood donation supply using a hybrid approach that combines Generative AI techniques with traditional machine learning algorithms. The model forecasts blood donation rates based on several Critical Success Factors (CSFs):

a. C1: State Population–The state's population size impact on the potential pool of blood donors.
b. C2: Day of the Week–Variations in donation rates on different days.
c. C3: Public Events/Holidays–Public events and holidays affect donation activities.
d. C4: Historical Donation Trends–Trends in past donations to predict future rates.
e. C5: Blood Group Distribution–Distribution of blood groups to ensure a balanced supply.

The methodology follows these steps:

a. Data Collection and Preprocessing: Data on the CSFs were gathered from sources like BBISv2, followed by preprocessing to ensure data consistency and accuracy.

b. Synthetic Data Generation: GANs and VAEs were used to generate synthetic data to supplement real-world data, enhancing model robustness.
c. Feature Engineering: CSFs were normalized and selected as features for the model, and PCC analysis was conducted to identify the most impactful factors.
d. Model Development: A Random Forest model was developed and trained on the combined real and synthetic datasets using CSFs as input features.
e. Model Evaluation and Iterative Improvement: The model's performance was assessed using metrics like accuracy, precision, recall, and AUC-ROC, followed by iterative refinement.
f. Final Prediction and Validation: The final model was validated against a test dataset to ensure reliability.

Figure 1 below illustrates the step-by-step process of developing a predictive model for blood donation rates. The flow begins with data collection and preprocessing, followed by synthetic data generation if necessary. Key features are selected through feature engineering, and a Random Forest model is developed and trained on the combined dataset. The model undergoes evaluation, with iterative improvements made as needed. The final prediction is validated against a test dataset, ensuring the model's accuracy and reliability.

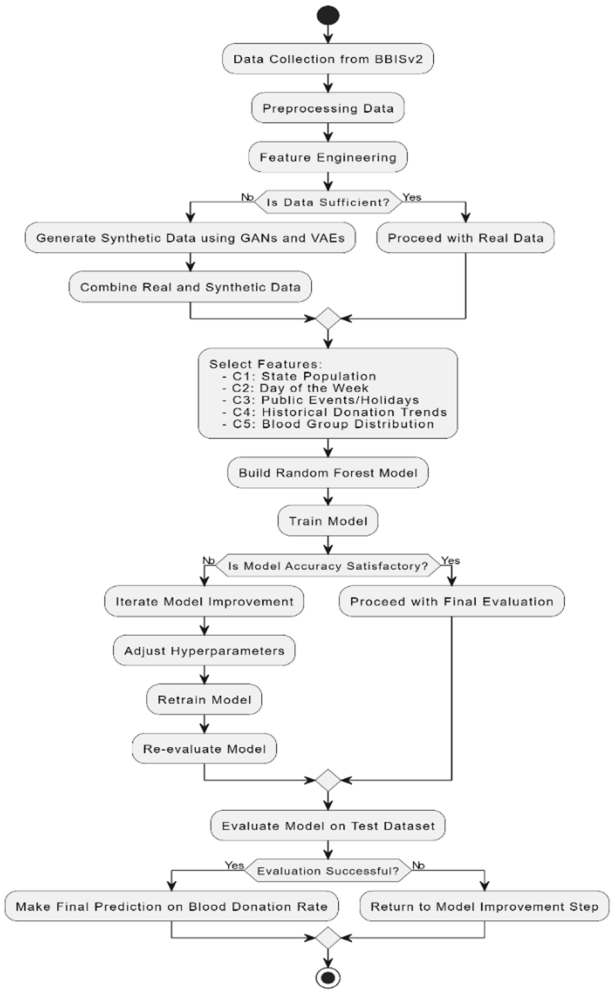

Fig. 1 Blood Donation Prediction Model Flow

B. Materials

This study utilized several key materials for developing and validating the predictive model:

1) Primary Dataset: The dataset from the Blood Bank Information System (BBISv2) covers approximately 80% of Malaysia's blood donation activities across 22 main collection sites. It provides detailed daily records, including donation dates, blood group types, state locations, and donation volumes. Rigorous preprocessing was conducted to address missing values, resolve inconsistencies, and normalize the data, ensuring a robust foundation for the model.

2) Synthetic Data Generation Tools: Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) were used to generate synthetic data, supplementing the primary dataset. These tools produced additional data points that mimic real-world donation patterns, enhancing the dataset's diversity and improving the model's robustness.

3) Computational Resources: High-performance computing resources, including GPU-enabled servers, facilitated the processing of complex algorithms in Generative AI and Random Forest modeling. These resources ensured efficient handling of the extensive dataset and the iterative training process.

4) Software and Libraries: Python, along with libraries like TensorFlow, Keras, Scikit-learn, and Pandas, was used for data preprocessing, synthetic data generation, model development, and performance evaluation. These tools provided the computational and analytical capabilities required for the predictive model.

C. Generative AI-Powered Predictive Analytics Model Development

This study employs GANs and VAEs to generate synthetic data, enhancing the predictive model's capacity to forecast blood donation patterns. By supplementing real data, these techniques address data sparsity, particularly in underrepresented regions and rare blood types, thereby improving the model's robustness and accuracy.

1) GANs Algorithm Operationalization: GANs involve two networks: a Generator and a Discriminator. The Generator c reates synthetic data mimicking real-world blood donation scenarios, while the Discriminator evaluates the data's authenticity. The training process is formulated as a min-max game between these two networks as per Equation (1):

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))] \quad (1)$$

where:

a. *pdata(x)* denotes the distribution of real blood donation data, such as daily donation counts by state and blood type. This data provided the ground truth for the model, including critical factors affecting blood donation supply.
b. *pz(z) represents* the distribution of input noise fed into the Generator. This noise served as a starting point for

the Generator to create synthetic data that attempts to mirror real-world blood donation scenarios.

c. *V(D, G)* is the *value* function that defines the overall objective of the GAN model. The Generator sought to minimise this value by producing synthetic blood donation data that could convincingly mimic real data, thereby deceiving the Discriminator. Conversely, the Discriminator aimed to maximise this value by accurately distinguishing between real blood donation data and the synthetic data generated by the model.

d. *D(x)* is the *Discriminator's* probability estimate that a given data instance *x* is real. The Discriminator was trained on both real data and synthetic data generated by the Generator. *D(x)* outputs a value between 0 and 1, where 1 indicates high confidence that *x* is real and 0 indicates confidence that *x* is synthetic.

e. *D(G(z)) represents* the Discriminator's probability estimate that the blood donation data generated by *G(z)* is real. The Generator aimed to create data such that $D(G(z)) \approx 1$, meaning that the Discriminator would find it difficult or impossible to distinguish the synthetic blood donation data from the real data. This capability was critical in ensuring that the synthetic data closely resembled real-world blood donation scenarios, thereby enhancing the overall robustness of the predictive model.

This adversarial framework refines the Generator's ability to produce realistic data, enhancing the dataset's diversity and realism. The synthetic data generated improves the model's accuracy by addressing data scarcity and simulating various potential blood donation outcomes.

2) VAEs Algorithm Operationalization: VAEs use an encoder-decoder architecture to generate synthetic blood donation data. The encoder maps input data into a latent space, and the decoder reconstructs this data, capturing the underlying distribution of the original dataset. The VAE's objective function combines reconstruction loss and a regularization term as per Equation (2):

$$\mathcal{L}(\theta, \phi; x) = \mathbb{E}_{q_\phi(z|x)}[\log\ p_\theta(x|z)] - \beta \cdot D_{KL}(q_\phi(z|x)||p(z)) \quad (2)$$

where:

a. $L(\theta,\phi;x)$ denotes the VAE's loss function for a specific blood donation data point x, parameterized by θ (decoder parameters) and ϕ (encoder parameters). This loss function is tailored to accurately capture the nuances of blood donation patterns across different states and blood types.

b. The first term, $\mathbb{E}_{q_\phi(z|x)}[log\ p_\theta(x|z)]$, represents the reconstruction loss, which measures how well the decoder can regenerate the original blood donation data from its encoded latent representation. This ensures that the synthetic data maintains fidelity to the real-world blood donation scenarios.

c. The second term, $D_{KL}(q_\phi(z|x)||p(z))$, is the Kullback-Leibler divergence, serving as a regularization mechanism. It ensures that the distribution of latent variables—reflective of the critical success factors influencing blood donation supply—remains aligned with a prior distribution. This regularization is crucial for preventing overfitting and ensuring that the generated data is generalizable to new, unseen scenarios.

d. The hyperparameter β balances the reconstruction accuracy against the regularization imperative. It was carefully tuned to ensure that the synthetic blood donation data generated by the VAEs maintained high fidelity to the complexities introduced by the CSFs identified in the study.

By employing VAEs in this manner, the study was able to produce a rich and varied dataset that enhanced the model's predictive capabilities, particularly in scenarios with limited real-world data. The synthetic data generated by VAEs allowed the predictive model to explore a broader range of potential blood donation outcomes, thereby improving its robustness and accuracy in forecasting blood donation supply across different regions and blood types in Malaysia.

3) Feature Engineering - Normalizing CSFs: Feature engineering is crucial for developing a predictive model, as it transforms raw data into useful features that enhance predictive accuracy. In this study, key Critical Success Factors (CSFs) influencing blood donation rates, including state population, day of the week, public events/holidays, historical donation trends, and blood group distribution, were identified. These CSFs were normalized to ensure consistent evaluation and comparison across different factors. Normalization is an integral part of the feature engineering process. It involves scaling each CSF to a uniform range of 1 to 10, allowing the model to analyze and compare the influence of diverse factors on blood donation rates equitably. This standardization ensures that the model can proportionally assess each factor's impact without being skewed by varying scales in the raw data. To ensure effective comparison and analysis, each Critical Success Factor (CSF) was normalized to a 1-10 scale as per Equation (3):

$$\text{Normalized Score}_\text{i} = \left(\frac{\text{Value}_\text{i} - \min(\text{Value})}{\max(\text{Value}) - \min(\text{Value})}\right) \times 9 + 1 \quad (3)$$

where:

a. Value_i represents the raw value of the CSF for a particular state.

b. min(Value)and max(Value) denote the minimum and maximum values of that CSF across all states, respectively.

This normalization ensures a uniform scale where 1 indicates the least favourable condition and 10 the most favourable. Examples include:

a. State Population (C1): Higher scores for larger populations.

b. Day of the Week (C2): Scores reflecting typical donation activity fluctuations.

c. Public Events/Holidays (C3): Scaled based on historical impact on donations.

d. Historical Donation Trends (C4): Higher scores for consistent or increasing trends.

e. Blood Group Distribution (C5): Scaled according to availability and demand.

Table I illustrates normalized data for five states. The "Normalized Donation Rate" represents the likelihood of

adequate blood supply, with higher scores indicating better conditions. This normalization ensures accurate prediction and equitable comparison across different states.

TABLE I
NORMALIZED DATA BASED ON CSFS FOR BLOOD DONATION SUPPLY

CSF	State A	State B	State C	State D	State E
C1: Population	8.28	3.15	9.96	2.57	3.29
C2: Day	8.77	3.10	10.00	2.67	2.51
C3: Events/Holidays	8.77	3.10	10.00	3.52	2.33
C4: Trends	8.39	3.14	9.97	2.19	3.32
C5: Blood Group Risk	8.36	3.12	9.98	4.34	2.74
Normalized Donation Rate	8.40	3.11	9.98	4.68	2.63

The Normalized Donation Rate is a crucial outcome metric in the predictive model. It represents the relative likelihood of a sufficient blood donation supply in each state, given the set of CSFs. Higher scores indicate states that are more likely to have adequate blood supplies, considering factors like population size, timing, public events, historical trends, and the distribution of blood groups.

4) Synthetic Data Generation: After normalizing the CSF, the next step involved enriching the dataset by generating synthetic data using GANs and VAEs. These models produced 50 synthetic datasets, simulating diverse blood donation scenarios across different states. Five random states (labelled as States A, B, C, D, and E) were selected to generate synthetic CSF data and corresponding normalized donation rates. By utilizing the normalized CSFs as a foundation, the synthetic data captured a broad spectrum of real-world conditions and variations. The purpose of employing GANs and VAEs was to expand the dataset, enhancing the model's generalizability. The 50 synthetic datasets reflected various potential blood donation scenarios, directly based on the normalized CSFs, thereby introducing a rich diversity of conditions that mirrored real-world patterns. This approach effectively addressed data scarcity issues and provided a more robust foundation for model training. Table II below presents a sample of the generated synthetic datasets. Each row represents a hypothetical state scenario with its normalized CSFs, offering a comprehensive view of potential real-world situations. The "Normalized Donation Rate (Prediction)" column indicates the predicted likelihood of adequate blood supply in each scenario.

TABLE II
SAMPLE OF SYNTHESIZED SYNTHETIC DATA FOR BLOOD DONATION SUPPLY

CSF	State A	State B	State C	State D	State E
C1: Population	7.2	3.1	10.0	4.6	2.3
C2: Day of the Week	5.3	3.4	9.8	6.5	4.1
C3: Events/Holidays	1.2	8.9	7.8	1.1	2.0
C4: Trends	8.9	4.3	7.2	3.5	7.9
C5: Blood Group	6.7	9.9	5.1	3.2	7.8
Normalized Donation Rate (Prediction)	8.1	3.4	9.0	4.1	6.2

5) Synthetic Data Validation: Using GANs and VAEs, a total of 50 synthetic datasets were generated, simulating blood donation scenarios across different states based on the identified CSFs. To ensure that the synthetic data was a reliable representation of real-world conditions, a rigorous validation process was undertaken:

a. Mean: The mean of each CSF in the synthetic data was calculated and compared to the mean of the real data to ensure consistency. The mean provides an average value of each CSF, helping to verify that the synthetic data aligns with the general trends observed in the real blood donation data as per Equation (4):

$$\text{Mean} = \frac{1}{n}\sum_{i=1}^{n} \quad x_i \tag{4}$$

where *xi* represents a CSF's individual values, such as a state's population or the frequency of blood donations on a specific day, and n is the total number of synthetic data points. This step ensures that the overall level of each CSF in the synthetic data mirrors that of the real data, maintaining the realistic distribution of factors that affect blood donation rates.

b. Variance: The variance was calculated to measure the spread of CSF values from the mean. This metric assesses how much the CSF values in the synthetic data differ from the average value, providing insights into the diversity and variability of the synthetic data compared to the real data as per Equation (5):

$$\frac{1}{n}\sum_{i=1}^{n} \quad (x_i - \text{Mean})^2 \tag{5}$$

In this context, variance helps ensure that the synthetic data captures the natural fluctuations in blood donation rates and other CSFs across different states and time periods.

c. Skewness: Skewness was calculated to determine the asymmetry of the CSF distribution in the synthetic data. This metric helps to identify whether the distribution of CSFs in the synthetic data is skewed towards higher or lower values compared to the real data, which could indicate a bias in the synthetic data generation process as per Equation (6):

$$\frac{1}{n}\sum_{i=1}^{n} \quad \left(\frac{x_i - \text{Mean}}{\text{Standard Deviation}}\right)^3 \tag{6}$$

By analyzing skewness, we can ensure that the synthetic data accurately reflects the natural skew present in real-world blood donation data, such as the tendency for more donations to occur during certain periods or in specific regions.

d. Kurtosis: Kurtosis was used to assess the peakedness of the CSF distribution in the synthetic data. This metric indicates whether the data distribution is more peaked or flatter than the real data distribution, which could affect the predictive model's ability to generalize from the synthetic data as per Equation (7):

$$\frac{1}{n}\sum_{i=1}^{n} \quad \left(\frac{x_i - \text{Mean}}{\text{Standard Deviation}}\right)^4 \tag{7}$$

Kurtosis validation ensures that the synthetic data appropriately reflects the concentration of blood donation events, whether they are clustered around certain values or spread out more evenly. The difference in the power of the exponent causes skewness to measure how far and in what direction values are distributed around the mean, while

kurtosis measures whether the data is heavy-tailed or light-tailed compared to a normal distribution.

e. K-S Test**:** The K-S test was applied to compare the distributions of the real and synthetic data to ensure they followed the same distribution. This test helps verify that the overall distribution of CSFs in the synthetic data is statistically like that of the real data, confirming that the synthetic data is a valid representation of real-world conditions as per Equation (8):

$$D_{n,m} = \sup_x |F_n(x) - F_m(x)| \quad (8)$$

where *Dn,m* is the K-S statistic, Fn(x) is the empirical distribution function of the real CSF data, and Fm(x) is the empirical distribution function of the synthetic CSF data. This comparison ensures the synthetic data can reliably simulate real blood donation scenarios.

The synthetic data served as an essential supplement to real-world data, particularly in cases where real data was sparse or inconsistent. By introducing a wide range of possible scenarios, the synthetic data enabled the predictive model to learn from a more diverse set of examples, thereby improving its accuracy and generalizability. The integration of synthetic datasets into the model's training process ensured that the model could effectively predict blood donation supply across various states and conditions. The synthetic data validation process ensured that the synthetic data was robust, accurate, and representative of real-world conditions. This process, combined with the extensive generation of diverse scenarios, contributed to the development of a reliable and effective predictive model for managing blood donation supply.

6) Deployment of Random Forest: Random Forest was deployed due to its robust ability to handle non-linear relationships and high-dimensional data in predicting blood donation rates. The algorithm builds multiple decision trees from bootstrapped samples of the dataset, and each tree is trained on a random subset of features. The prediction for the blood donation rate, y^\hat{y}y^, is calculated as the average of the predictions from all trees in the forest, as shown in Equation (9):

$$\hat{y} = \frac{1}{N}\sum_{i=1}^{N} \quad T_i(x) \quad (9)$$

where $T_i(x)$ represents the prediction of the *i-th* decision tree, and x is the set of input features, including the normalized CSFs such as state population, public events, and blood group distribution. This aggregation of predictions ensures stable and accurate results, reducing overfitting.

7) Deployment of SVM: SVM was used to classify blood donation scenarios as either sufficient or insufficient based on the CSFs. The SVM seeks to find the optimal hyperplane that maximizes the margin between the two classes. The optimization problem for SVM is expressed in Equation (10):

$$\min_{\mathbf{w},b} \frac{1}{2}|\mathbf{w}|^2$$
$$\text{subject to } y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 \forall i \quad (10)$$

where w is the weight vector, b is the bias, *xi* represents the CSFs for *the i-th* data point, and *yi* is the label (sufficient or insufficient blood donation rate). Both linear and non-linear kernels were evaluated to determine the best fit for capturing blood donation trends.

8) Deployment of Neural Networks: A feedforward Neural Network with multiple hidden layers was used to model non-linear relationships between CSFs and blood donation rates. Each neuron computes a weighted sum of the input features, as represented by Equation (11):

$$w_{t+1} = w_t - \eta\frac{1}{2}\sum_{i=1}^{n} \quad (y_i - a(\mathbf{w}^T\mathbf{x} + b))^2\text{T} \quad (11)$$

In this context, w_t and w_{t+1} represent the weights before and after the update, respectively, while η denotes the learning rate, controlling the size of the weight adjustments during training. The actual donation rate is denoted by *yi,* and the activation function *a(z)* can be either the sigmoid function (which maps the output between 0 and 1) or the Rectified Linear Unit (ReLU) function (which outputs 0 for negative values and passes through positive values).The term $(\mathrm{w}^T\mathrm{x} + b)$ represents the weighted sum of the input features, where w is the weight vector, *x* is the input, and b is the bias. This process enables the neural network to progressively learn from the data, minimizing the error between the predicted donation rates and the actual donation rates.

9) Deployment of Linear Regression: Linear Regression was deployed to establish a linear relationship between CSFs and blood donation rates. The model fits a linear equation to predict the donation rate y, represented by Equation (12):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \quad (12)$$

where $x_1, x_2 \ldots$represent the CSFs, and *β0, β1…βn* are the coefficients learned during training. Although simple, Linear Regression serves as a useful baseline for comparison with more complex models.

10) Predictive Model Training: Training involved integrating real-world data with the 50 synthetic datasets generated by GANs and VAEs to address data sparsity. The Random Forest algorithm was chosen for its robustness in handling complex interactions. Feature selection was guided by the PCC, identifying the most relevant CSFs to reduce complexity and improve accuracy. Hyperparameters were fine-tuned, and cross-validation was employed to prevent overfitting. The inclusion of synthetic data allowed the model to handle edge cases and outliers effectively. This rigorous training approach resulted in a reliable predictive tool capable of optimizing blood resource allocation across Malaysia.

11) Continuous Monitoring and Adaptation of the Model: To ensure long-term effectiveness, the predictive model for blood donation supply undergoes continuous monitoring and adaptation. Advanced analytics dashboards track key metrics such as accuracy, precision, recall, and AUC-ROC in real time. Any performance degradation is promptly addressed through automated retraining pipelines that incorporate new data and user feedback, allowing the model to adapt to changing donation patterns. The retraining process can be mathematically represented as per Equation (13):

$$\theta t + 1 = \theta t - \eta \nabla L(\theta t; Dt) \quad (13)$$

In predicting blood donation supply, the formula provided plays a critical role in continuously updating and refining the model to ensure its predictions remain accurate and relevant. Here's how the parameters are explained in this setting:

a. θt represents the model parameters at iteration. In the case of blood donation prediction, these parameters could include the weights assigned to various factors such as state population, day of the week, public events, historical donation trends, and blood group distribution. These parameters are adjusted at each iteration to improve the model's accuracy.
b. η is the learning rate, which controls how much the model parameters are adjusted in response to new data. For example, suppose new data indicates a significant change in blood donation patterns (such as a sudden increase in donations due to a public awareness campaign). In that case, the learning rate determines how quickly the model adapts to this new information. A higher learning rate results in larger adjustments, while a lower rate leads to more gradual changes.
c. $\nabla L(\theta t; Dt)$ is the gradient of the loss function L concerning the model parameters. The loss function L measures the difference between the predicted and actual blood donation rates observed in the data. The gradient provides direction on how to adjust the parameters θt to reduce this difference. Minimizing the loss function makes the model more accurate in predicting blood donation rates.

By periodically integrating new datasets and leveraging machine learning techniques like reinforcement learning, the model adapts to real-time changes in blood donation patterns. Additionally, synthetic data generated by GANs and VAEs is periodically updated to reflect evolving trends, ensuring the model remains robust and accurate.

12) Development of the Predictive Analytics Model Interface Using Python: The Blood Donation Predictive Analytics Model was developed using Python, employing libraries such as Tkinter, Scikit-learn, TensorFlow, Keras, Pandas, and NumPy. The Tkinter-based interface enables users to input various Critical Success Factors (CSFs) affecting blood donation rates, including state population, day of the week, public events, historical trends, and blood group distribution.

The model outputs a normalized donation rate on a scale from 0 to 10:

a. 0-3: Low supply, indicating a potential shortfall and the need for immediate action like donation campaigns.
b. 4-7: Moderate supply, suggesting regular drives and trend monitoring to prevent shortages.
c. 8-10: Healthy supply allows resource reallocation as needed.

This user-friendly interface provides healthcare professionals and policymakers quick, actionable insights to optimize blood resource management and support informed decision-making in real-time scenarios. As shown in Figure 2.0 below, the prototype interface allows users to input various CSFs influencing blood donation rates, such as state population, day of the week, public events/holidays, historical donation trends, and blood group distribution.

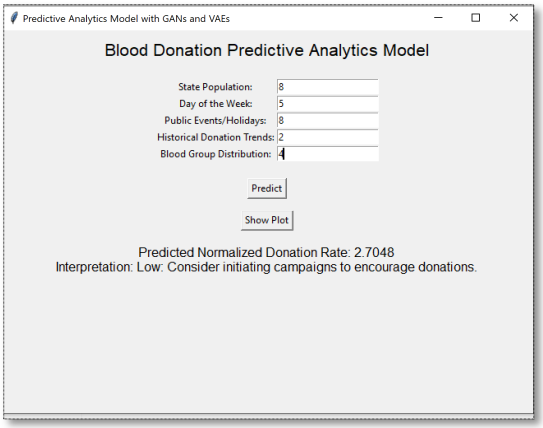

Fig. 2 Blood Donation Predictive Analytics Model Interface

D. Model Evaluation of Predictive Accuracy: The Quantitative Approach

The model's predictive accuracy was evaluated using a set of performance metrics to ensure its reliability in forecasting blood donation supply across different states and blood types. The following metrics were used to assess the model's effectiveness: It is calculated as per Equation (14):

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{14}$$

where TP represents true positives (correctly predicted adequate blood supply), TN represents true negatives (correctly predicted inadequate blood supply), FP represents false positives (incorrectly predicted adequate blood supply), and FN represents false negatives (incorrectly predicted inadequate blood supply). Precision evaluates the proportion of true positives out of the total predicted positives, indicating the model's accuracy in predicting days with sufficient blood supply. It is calculated as per Equation (15):

$$\text{Precision} = \frac{TP}{TP+FP} \tag{15}$$

Recall (Sensitivity): Recall measures the proportion of true positives out of the actual positives, assessing the model's ability to identify all days with sufficient blood supply. It is calculated as per Equation (16):

$$\text{Recall} = \frac{TP}{TP+FN} \tag{16}$$

The AUC-ROC comprehensively evaluates the model's ability to discriminate between days with adequate and inadequate blood supply. The ROC curve plots the true positive rate (recall) against the false positive rate (FPR), which is calculated as per Equation (17):

$$\text{FPR} = \frac{FP}{FP+TN} \tag{17}$$

This set of metrics ensured a thorough comparison of different predictive models, verifying that the chosen model could reliably predict variations in blood donation supply, particularly across different blood types and states.

III. Results and Discussion

This section presents the findings from evaluating various machine learning models developed for forecasting blood donation supply. The models, including Random Forest, SVM, Neural Networks, and Linear Regression, were trained

using synthetic datasets generated via GANs and VAEs The synthetic data was rigorously validated, demonstrating high accuracy and minimal deviations from real-world data across key statistical metrics. The evaluation aimed to assess the predictive performance of these models in forecasting blood donation outcomes.

A. Validation of Generative AI Model: Ensuring Synthetic Data Accuracy for Blood Donation Predictions

The synthetic data for blood donation supply, consisting of 50 datasets generated by GANs and VAEs, was validated to ensure its accuracy in reflecting real-world blood donation patterns. This validation involved comparing summary statistics—mean, variance, skewness, and kurtosis—between the real and synthetic data, and conducting a K-S test to verify the distribution alignment.

TABLE III
MEAN COMPARISON

CSF	Real Data Mean	Synthetic Data Mean
C1: State Population	5.44	5.45
C2: Day of the Week	5.42	5.45
C3: Public Events/Holidays	4.2	4.54
C4: Historical Donation Trends	6.36	5.86
C5: Blood Group Distribution	6.54	6.44
Normalized Donation Rate	5.44	5.45

Table III above shows that the mean values for the synthetic data closely mirror the real data, with no more than 0.34 differences across all CSFs. This tight alignment indicates that the synthetic data accurately captures the average behavior of key factors influencing blood donation. With deviations under 10%, the synthetic data can be confidently used for predictive modeling without introducing significant bias or misrepresenting real-world tendencies.

TABLE IV
VARIANCE COMPARISON

CSF	Real Data Variance	Synthetic Data Variance
C1: State Population	7.9975	9.3239
C2: Day of the Week	8.1824	9.6209
C3: Public Events/Holidays	13.524	15.041
C4: Historical Donation Trends	7.9875	9.8739
C5: Blood Group Distribution	10.357	12.457
Normalized Donation Rate	9.4124	10.532

Table IV highlights a slight increase in variance in the synthetic data compared to the real data, with differences ranging from 1 to 2 units across CSFs. The higher variance in the synthetic data suggests a broader representation of possible outcomes, which can be advantageous for generating more flexible predictions. While slightly more spread out, the synthetic data still effectively captures the variability present in real-world donation behaviors, enhancing the model's ability to forecast diverse scenarios.

TABLE V
SKEWNESS COMPARISON

CSF	Real Data Skewness	Synthetic Data Skewness
C1: State Population	0.48	0.5
C2: Day of the Week	0.35	0.38
C3: Public Events/Holidays	0.52	0.55
C4: Historical Donation Trends	0.49	0.51
C5: Blood Group Distribution	0.42	0.45
Normalized Donation Rate	0.48	0.5

Table V demonstrates that the synthetic data closely matches the skewness of the real data, with differences in skewness values between 0.02 and 0.03 across all CSFs. This indicates that the synthetic data replicates the asymmetry of the real data's distribution, ensuring that both common and less frequent donation behaviors are accurately represented. Such alignment in skewness is essential for making reliable predictions about irregular or atypical donation patterns.

TABLE VI
KURTOSIS COMPARISON

CSF	Real Data Kurtosis	Synthetic Data Kurtosis
C1: State Population	1.77	1.8
C2: Day of the Week	1.64	1.68
C3: Public Events/Holidays	2.2	2.26
C4: Historical Donation Trends	1.91	1.96
C5: Blood Group Distribution	1.82	1.85
Normalized Donation Rate	2.05	2.1

Table VI illustrates that the kurtosis values for the synthetic data are nearly identical to those of the real data. Differences are minimal, remaining within a 0.05 range across all CSFs. This indicates that the synthetic data captures the occurrence of outliers and extreme values with high accuracy, which is crucial for modeling rare but impactful donation events. The similar kurtosis values validate the synthetic data's capacity to reflect real-world donation spikes, enhancing the robustness of the predictive model.

TABLE VII
K-S TEST RESULTS

CSF	Dn,m
C1: State Population	0.2
C2: Day of the Week	0.4
C3: Public Events/Holidays	0.45
C4: Historical Donation Trends	0.4
C5: Blood Group Distribution	0.35
Normalized Donation Rate	0.3

Table VII lists the K-S test results, showing the maximum distance between the empirical distribution functions of the real and synthetic data. The K-S test results in Table VII confirm that the synthetic data closely follows the distribution of the real data, with Dn,m values below 0.5 across all CSFs. For the Normalized Donation Rate, the Dn,m value of 0.30

shows a powerful alignment between the real and synthetic data. This ensures that the synthetic data effectively captures the overall distribution trends, validating its use in forecasting real-world blood donation outcomes confidently.

B. Quantitative Comparative Study: Performance Evaluation of Machine Learning Models

The predictive performance of various machine learning models was analyzed to assess their efficacy in forecasting blood donation rates. The models evaluated include Random Forest, SVM, Neural Networks, and Linear Regression. All models were trained and tested on the same set of 50 synthetic datasets generated using CSFs to ensure consistency and integrity across comparisons. The same CSF input was applied to each model to generate predictions, which were then compared against real prediction data to validate performance. The models' performance was assessed using key metrics such as accuracy, precision, recall, and AUC-ROC.

TABLE VIII
SUMMARY OF MODEL PERFORMANCE COMPARISONS

Model Type	Predicted Donation Rate	Accuracy	Precision	Recall	AUC-ROC
Random Forest	8.4084	98.70%	0.91	0.89	0.92
Neural Networks	7.5337	87.60%	0.72	0.7	0.74
SVM	8.1583	93.20%	0.81	0.79	0.83
Linear Regression	8.3975	98.60%	0.89	0.88	0.9

Table VIII above presents a comparative summary of the models' predicted donation rates alongside the key performance metrics. The results indicate that the Random Forest model demonstrated the highest accuracy (98.7%) and superior precision (0.91), followed closely by the Linear Regression model, which achieved an accuracy of 98.6%. Both models produced predictions that were very close to the actual donation rate, with minimal deviation. The Neural Networks model exhibited the lowest performance across all metrics, suggesting that its predictive capabilities were less effective for this dataset. The SVM model performed moderately, with an accuracy of 93.2%, but did not match the predictive strength of Random Forest or Linear Regression.

This study evaluates the predictive performance of four machine learning models—Random Forest, SVM, Neural Networks, and Linear Regression—in forecasting blood donation rates. By leveraging 50 synthetic datasets generated from CSFs all models were trained and tested under the same conditions, ensuring consistency and comparability. Using validated synthetic data enabled a robust framework for assessing model performance against real-world data, providing theoretical and practical implications for blood donation forecasting. From a theoretical perspective, this study demonstrates the importance of model selection in predictive analytics, particularly for time-sensitive tasks such as blood donation supply management. The findings reinforce that Random Forest, as an ensemble learning method, outperforms other traditional models in handling high-dimensional data and capturing complex interactions between features. Its decision tree-based structure, combined with feature bagging, makes it particularly effective in scenarios where multiple factors (e.g., population, public events, holidays) contribute to the prediction outcome. The superior accuracy (98.7%), precision (0.91), and AUC-ROC (0.92) of the Random Forest model highlight its robustness in real-world applications, as it consistently outperformed other models across various metrics.

Additionally, the results indicate that Linear Regression, despite being a simpler algorithm, showed strong predictive power (accuracy of 98.6%). This aligns with theoretical expectations that linear models can effectively capture relationships when the input-output behavior follows a linear trend. However, its limitations become apparent when dealing with more complex, nonlinear data structures, where models like Random Forest excel. In contrast, Neural Networks, often regarded for their ability to model non-linear relationships, underperformed in this context. The lower accuracy (87.6%) and precision (0.72) suggest that the neural network model struggled with the dataset's characteristics, possibly due to overfitting or insufficient tuning of hyperparameters. This highlights an important theoretical consideration: while neural networks are powerful in complex scenarios, they may not always outperform simpler models when the underlying data does not warrant such complexity. SVM also exhibited moderate performance (accuracy of 93.2%), which is consistent with the model's sensitivity to the choice of kernel function and regularization parameters, factors that may have limited its performance in this study.

The practical implications of these findings are significant for healthcare systems, particularly in the context of blood donation supply management. Accurate forecasting of donation rates is crucial for optimizing resource allocation, scheduling donation drives, and maintaining adequate blood supplies. The superior performance of the Random Forest model suggests that it can be reliably deployed to predict donation trends, especially in data-rich environments where multiple factors influence donation behavior. By incorporating Random Forest into decision-making frameworks, healthcare providers can reduce the risk of shortages, improving patient outcomes and operational efficiency.

Moreover, validating synthetic data through multiple statistical tests (mean, variance, skewness, kurtosis, and the K-S test) emphasizes the value of synthetic data generation for real-world applications. When real data is scarce, sensitive, or incomplete, using high-quality synthetic datasets, as demonstrated in this study, offers a practical solution for training machine learning models. In this context, the small deviations (less than 10%) between real and synthetic data regarding mean and distribution metrics provide confidence that synthetic data can be reliably used for future forecasting. Adopting predictive models like Random Forest, trained on validated synthetic data for blood banks and public health organizations can support more efficient planning and real-time decision-making. For instance, during high-demand periods (e.g., holidays or public events), these models can be integrated into automated systems that provide actionable insights, enabling rapid adjustments to donation strategies. This study also contributes to the broader discussion on the applicability of machine learning in healthcare. By illustrating the trade-offs between different models, this research highlights that more complex models, such as Neural

Networks, may not always yield better results in specific domains. Instead, simpler models, such as Linear Regression, and ensemble methods, like Random Forest, can outperform neural networks when data complexity is moderate, as with blood donation trends.

Additionally, the findings emphasize the importance of data quality in predictive modeling. The accurate and consistent performance of models trained on high-quality synthetic data underscores the potential for Generative AI techniques (such as GANs and VAEs) to mitigate challenges related to data scarcity or incomplete datasets. As healthcare systems continue to digitize and collect vast amounts of data, the ability to generate and use synthetic data will play an increasingly important role in training models that generalize well in real-world scenarios.

IV. Conclusion

This study successfully achieved its research objectives by evaluating the performance of various machine learning models in forecasting blood donation rates. Using synthetic datasets generated via Generative AI techniques such as GANs and VAEs, the study ensured that the models were trained on high-quality, validated data that closely mirrored real-world donation trends. The statistical validation of the synthetic data showed minimal deviations in mean, variance, skewness, kurtosis, and distribution alignment (K-S test), confirming that the synthetic data accurately represented real-world behaviours, with deviations under 10%. The findings demonstrated that the Random Forest model outperformed other models, achieving the highest accuracy (98.7%), precision (0.91), and AUC-ROC (0.92), making it the most reliable for predicting blood donation rates. Linear Regression also showed strong predictive performance, with an accuracy of 98.6%, proving to be a robust alternative to Random Forest. In contrast, the Neural Networks and SVM models underperformed, with lower accuracy and precision, suggesting they were less suited to the specific characteristics of the dataset used in this study. The validation of the synthetic data used for training these models is particularly noteworthy, as it ensures that the models' predictions are reliable and applicable in real-world scenarios. The slight increase in variance in the synthetic data compared to the real data provided a broader scenario representation, further enhancing the models' generalizability.

This study demonstrates the efficacy of machine learning models, particularly Random Forest, to predict blood donation rates, ensuring more informed decision-making and better resource allocation in healthcare management. The success of using validated synthetic data provides a strong foundation for future predictive analytics efforts in healthcare, where real-world data may be limited or unavailable. The findings contribute to the field by emphasizing the role of high-quality synthetic data in enhancing the accuracy and reliability of machine learning models for critical forecasting tasks, such as managing blood donation supply.

Acknowledgment

This study benefited from the valuable contributions of several individuals. Special thanks are due to Abdul Samad Bin Shibghatullah from UCSI University Kuala Lumpur for his technical expertise in data analysis and methodology, and to Chloe Thong Chee Ling for her insights on data modeling. Nazirul Nazrin Bin Shahrol Nidzam from Universiti Tenaga Nasional provided critical feedback on data validation, while Samer Muthana Sarsam from Coventry University offered valuable perspectives on strategic direction. Appreciation is also extended to Shayla Islam and Raenu AL Kolandaisamy for their guidance and manuscript review. Additionally, Halimah Badioze Zaman, Chair Holder of the Tan Sri Leo Moggie Distinguished Chair in Energy Informatics at UNITEN, is acknowledged for her advice and support throughout the research process. This study was conducted without external funding from public, commercial, or not-for-profit sectors. The findings, interpretations, and conclusions expressed are solely those of the authors and do not necessarily reflect the views of their affiliated institutions

References

[1] R. Sethuraman, S. Murugan, and M. Saravanan, "AI and Machine Learning in Supply Chain Optimization: Mapping the Territory," *Blockchain, IoT, and AI Technologies for Supply Chain Management*, pp. 177–220, 2024, doi: 10.1007/979-8-8688-0315-4_7.

[2] P. Maheshwari, S. S. Kamble, A. Belhadi, S. Gupta, and S. K. Mangla, "Resilient healthcare network for simultaneous product allocations during supply chain disruptions," *Supply Chain Forum: An International Journal*, vol. 25, no. 4, pp. 407–427, Aug. 2023, doi:10.1080/16258312.2023.2238669.

[3] J. D. VanVactor, "Healthcare supply chain resiliency," *Journal of Supply Chain Management, Logistics and Procurement*, vol. 3, no. 2, p. 148, Dec. 2020, doi: 10.69554/kugd8447.

[4] K. A. Clauson, E. A. Breeden, C. Davidson, and T. K. Mackey, "Leveraging Blockchain Technology to Enhance Supply Chain Management in Healthcare:," *Blockchain in Healthcare Today*, Mar. 2018, doi: 10.30953/bhty.v1.20.

[5] B. Scala and C. F. Lindsay, "Supply chain resilience during pandemic disruption: evidence from healthcare," *Supply Chain Management: An International Journal*, vol. 26, no. 6, pp. 672–688, May 2021, doi:10.1108/scm-09-2020-0434.

[6] S.-W. Lin, K. N. Anisa, and Y.-C. Chen, "Effects of meteorological factors on blood donation," *Transfusion and Apheresis Science*, vol. 59, no. 6, p. 102901, Dec. 2020, doi: 10.1016/j.transci.2020.102901.

[7] K. Intharanut, P. Nimnuch, W. Naiwijarn, W. Pimsiri, O. Khantisitthiporn, and O. Nathalang, "Characteristics of Donors and Modelling of the Characteristics to Possible Forecast the Repeat Donors Profile at Thammasat University Hospital*," Indian Journal of Hematology and Blood Transfusion*, vol. 39, no. 1, pp. 146–150, Aug. 2022, doi: 10.1007/s12288-022-01552-y.

[8] T. Ahamed, Md. Nazrul Islam, S. M. Taslim Uddin Raju, and M. M. A. Hashem, "Blood Donor Arrival Forecasting Using Regression Model and Analysis of Donor Behavioural Pattern," *2022 25th International Conference on Computer and Information Technology (ICCIT)*, pp. 897–902, Dec. 2022, doi:10.1109/iccit57492.2022.10054710.

[9] H. Shih and S. Rajendran, "Comparison of Time Series Methods and Machine Learning Algorithms for Forecasting Taiwan Blood Services Foundation's Blood Supply," *Journal of Healthcare Engineering*, vol. 2019, pp. 1–6, Sep. 2019, doi: 10.1155/2019/6123745.

[10] A. Thijsen, C. N. Gemelli, T. E. Davison, and B. Masser, "Using the Health Action Process Approach to predict blood donation intentions and return behavior following a vasovagal reaction for whole blood and plasma donors," *Transfusion*, vol. 62, no. 9, pp. 1791–1798, Aug. 2022, doi: 10.1111/trf.17052.

[11] A. Adhikari, R. Joshi, and S. Basu, "Collaboration and coordination strategies for a multi-level AI-enabled healthcare supply chain under disaster," *International Journal of Production Research*, vol. 63, no. 2, pp. 497–523, Sep. 2023, doi: 10.1080/00207543.2023.2252933.

[12] A. Kumar, V. Mani, V. Jain, H. Gupta, and V. G. Venkatesh, "Managing healthcare supply chain through artificial intelligence (AI): A study of critical success factors," *Computers & Industrial Engineering*, vol. 175, p. 108815, Jan. 2023, doi:10.1016/j.cie.2022.108815.

[13] N. A. Alhamda and A. Z. Rahman, “The Impact of IoT Supply Chain Integrated Blockchain Technology Advancements in the Healthcare Sector: A Literature Review,” *Journal of Logistics and Supply Chain*, vol. 2, no. 1, pp. 33–40, Apr. 2022, doi: 10.17509/jlsc.v2i1.62835.

[14] P. Yu, H. Xu, X. Hu, and C. Deng, “Leveraging Generative AI and Large Language Models: A Comprehensive Roadmap for Healthcare Integration,” *Healthcare*, vol. 11, no. 20, p. 2776, Oct. 2023, doi:10.3390/healthcare11202776.

[15] A. Chansamut, “A Digital System Model for Healthcare Management in Thai Supply Chain,” *International Journal of Supply Chain Management*, vol. 12, no. 2, pp. 33–36, Apr. 2023, doi:10.59160/ijscm.v12i2.6127.

[16] S. Reddy, “Generative AI in healthcare: an implementation science informed translational path on application, integration and governance,” *Implementation Science*, vol. 19, no. 1, Mar. 2024, doi:10.1186/s13012-024-01357-9.

[17] M. Krishnan and A. Khastgir, “Enhancing Supply Chain Efficiency to Build Next-Gen Artificial Intelligence (AI)/Machine Learning Network Through Al-Driven Forecasting,” *International Journal of Supply Chain Management*, vol. 13, no. 3, pp. 28–41, Jun. 2024, doi:10.59160/ijscm.v13i3.6244.

[18] A. Longaray, N. Marube, L. Ensslin, A. Dutra, and S. Ensslin, “Performance evaluation of public healthcare supply chain management: a critical literature review,” *International Journal of Healthcare Technology and Management*, vol. 1, no. 1, p. 1, 2022, doi:10.1504/ijhtm.2022.10054715.

[19] J. Haw, K. Holloway, and M. Goldman, “How do we forecast tomorrow’s transfusion? Applying social science approaches to meet tomorrow’s transfusion needs: Blood donors and donation,” *Transfusion Clinique et Biologique*, vol. 30, no. 1, pp. 47–51, Feb. 2023, doi: 10.1016/j.tracli.2022.11.001.

[20] S. Pahune and N. Rewatkar, “Healthcare: A Growing Role for Large Language Models and Generative AI,” *International Journal for Research in Applied Science and Engineering Technology*, vol. 11, no. 8, pp. 2288–2301, Aug. 2023, doi: 10.22214/ijraset.2023.55573.

[21] T. Bou Assi, A. Haddad, L. Haddad, and O. Garraud, “Can a decentralized blood supply system reach 100% voluntary nonremunerated donation?,” *The International Journal of Health Planning and Management*, vol. 33, no. 4, Jul. 2018, doi:10.1002/hpm.2576.

[22] S. K. Thakur, “Forecasting demand for blood products: Towards inventory management of a perishable product,” *Bioinformation*, vol. 20, no. 1, pp. 20–28, Jan. 2024, doi: 10.6026/973206300200020.

[23] M. D. Wood, “Director General Joint Supply Chain, Defence Equipment and Support, 2003–08,” Who’s Who, Dec. 2007, doi:10.1093/ww/9780199540884.013.10000276.

[24] A. K. Gupta, G. V. Awatade, S. S. Padole, and Y. S. Choudhari, “Digital Supply Chain Management Using AI, ML and Blockchain,” *Innovative Supply Chain Management via Digitalization and Artificial Intelligence*, pp. 1–19, 2022, doi: 10.1007/978-981-19-0240-6_1.

[25] A. J. Scriffignano, “From supply chain to integrated value chain : Re-envisioning counterparty relationships in the context of AI and the convergence of disruption,” *Journal of Supply Chain Management, Logistics and Procurement*, vol. 1, no. 4, p. 334, Jun. 2019, doi:10.69554/xbnr3835.

[26] K. I. Afrashtehfar and A. S. H. Abu-Fanas, “Metaverse, Crypto, and NFTs in Dentistry,” *Education Sciences*, vol. 12, no. 8, p. 538, Aug. 2022, doi: 10.3390/educsci12080538.

[27] J. Nsikan, E. A. Affiah, I. Briggs, and N. Koko, “Sustainable supplier selection factors and supply chain performance in the Nigerian healthcare industry,” *Journal of Transport and Supply Chain Management*, vol. 16, Apr. 2022, doi: 10.4102/jtscm.v16i0.633.

[28] N. Labib and S. Gharib, “Healthcare in the Era of Generative AI,” *Industry 5.0 for Smart Healthcare Technologies*, pp. 265–274, Jul. 2024, doi: 10.1201/9781032632223-22.

[29] M. Patel Rohit Malik, “The Disruptive Influence of Generative AI in Life Science and Healthcare,” *International Journal of Science and Research (IJSR)*, vol. 13, no. 4, pp. 216–219, Apr. 2024, doi: 10.21275/sr24402184343.

[30] T. Hashizume *et al.*, "Development and validation of a scoring system to predict vasovagal reaction upon whole blood donation," *Vox Sanguinis*, vol. 119, no. 4, pp. 300–307, Dec. 2023, doi:10.1111/vox.13579.

[31] N. Al-Ma’aitah, “Investigating the interplay between supply chain agility, human capital and supply chain performance in the healthcare sector of Jordan,” *Uncertain Supply Chain Management*, vol. 12, no. 2, pp. 751–760, 2024, doi: 10.5267/j.uscm.2024.1.011.

[32] S. Selvaraj, “Empowering Patients with AI-Driven Personalized Care: The Transformative Power of Generative AI and Healthcare Data Integration,” *International Journal of Science and Research (IJSR)*, vol. 13, no. 7, pp. 337–343, Jul. 2024, doi: 10.21275/sr24703063340.

[33] R. Malik and K. Naudiyal, “Enabling Generative AI for Life Sciences and Healthcare Customers using the Power of Cloud,” *International Journal of Science and Research (IJSR)*, vol. 12, no. 11, pp. 1356–1360, Nov. 2023, doi: 10.21275/sr231115115845.

[34] O. Lemon, “Introducing generative AI into healthcare practice,” *The Journal for The Foundation of Science and Technology*, vol. 23, no. 7, Mar. 2024, doi: 10.53289/ovfa7327.

[35] S. Y. W. Chai, F. J. F. Phang, L. S. Yeo, L. H. Ngu, and B. S. How, “Future era of techno-economic analysis: Insights from review,” *Frontiers in Sustainability*, vol. 3, Aug. 2022, doi:10.3389/frsus.2022.924047.

[36] V. Yandrapalli, “Revolutionizing Supply Chains Using Power of Generative AI,” *International Journal of Research Publication and Reviews*, vol. 4, no. 12, pp. 1556–1562, Dec. 2023, doi:10.55248/gengpi.4.1223.123417.

[37] J. C. Essila, “Strategies for reducing healthcare supply chain inventory costs,” *Benchmarking: An International Journal*, vol. 30, no. 8, pp. 2655–2669, Aug. 2022, doi: 10.1108/bij-11-2021-0680.

[38] K. Lee, A. F. Cooper, and J. Grimmelmann, “Talkin’ ‘Bout AI Generation: Copyright and the Generative AI Supply Chain,” *SSRN Electronic Journal*, 2023, doi: 10.2139/ssrn.4523551.

[39] A. Reena Dhan and B. Kumar, “Machine Learning for Healthcare: Predictive Analytics and Personalized Medicine,” *International Journal of Science and Research (IJSR)*, vol. 13, no. 6, pp. 1307–1313, Jun. 2024, doi: 10.21275/mr24608013906.

[40] Y. R. Venugopal and S. V., "Ensemble Learning Approaches for Improved Predictive Analytics in Healthcare," *International Journal of Research Publication and Reviews*, vol. 5, no. 3, pp. 757–760, Mar. 2024, doi: 10.55248/gengpi.5.0324.0629.