e-ISSN: 3023-8609
RESEARCH ARTICLE
Vol.3 / No.1 / 2026 / pp.60-76
https://doi.org/10.69882/adba.cem.2026018

Enhancing Hospital Inventory Forecasting Accuracy through Hybrid and Ensemble Learning Models

Cem Özkurt[*,1], **Ahmet Kutey Küçükler**[α,2], **Murat Karslıoğlu**[β,3] **and Ruveyda Nur Özdemir**[§,4]
[*]Department of Computer Engineering, Sakarya University of Applied Sciences, Sakarya, Türkiye, [α]Department of Computer Engineering, Sakarya University of Applied Sciences, Sakarya, Türkiye, [β]Department of Computer Engineering, Sakarya University of Applied Sciences, Sakarya, Türkiye, [§]Department of Computer Engineering, Sakarya University of Applied Sciences, Sakarya, Türkiye.

ABSTRACT Demand forecasting for medical and consumable supplies in healthcare institutions is a challenging problem due to irregular usage patterns, seasonality, sudden demand spikes, and data sparsity, and inaccurate forecasts may lead to stock-outs, excessive inventory costs, and disruptions in patient care. This study proposes an anomaly-aware, hybrid and ensemble-based forecasting and decision support framework for short-term hospital inventory demand prediction using real-world operational data obtained from a hospital inventory management system. The proposed approach integrates density-based anomaly detection, material-level behavioral feature extraction, supervised time series transformation, and a multi-model ensemble architecture combining linear models, tree-based methods, and boosting-based learners, with model selection and weighting performed via time-series cross-validation. To ensure operational robustness, a multi-layer fallback strategy incorporating classical exponential smoothing and conservative heuristics is employed for data-scarce scenarios, and an interpretable rule-based forecast confidence score together with an integrated ABC–XYZ segmentation scheme is used to directly link forecasts with inventory control policies. Experimental results on real hospital inventory data demonstrate that the proposed framework significantly improves forecasting stability and accuracy compared to single-model approaches, particularly for heterogeneous and irregular consumption patterns, while providing a practical, explainable, and operationally actionable solution for hospital inventory management.

KEYWORDS
Hospital inventory management
Demand forecasting
Ensemble learning
Hybrid models
Time series forecasting
Anomaly detection
Decision support systems

INTRODUCTION

The effective and efficient operation of healthcare delivery systems plays a critical role in today's complex and dynamic healthcare environment. In hospitals, having the right amount of medical and consumable supplies in the right place at the right time is essential for the uninterrupted delivery of patient care services (Joshi *et al.* 2025). Inventory management is an important component not only in terms of operational efficiency, but also in terms of financial performance, the quality of hospital services, and human health. Incorrectly estimating the demand for medical and consumable supplies can have very serious consequences for hospitals. Excessive stockpiling leads to increased storage costs, deterioration of material quality, and significant financial losses, while insufficient stock levels cause disruptions in patient care, reduced treatment quality, and, in some cases, pose a threat to patient safety. Therefore, demand forecast accuracy is a fundamental requirement for optimizing hospital resources and delivering sustainable healthcare services (Umoren *et al.* 2025).

In previous years, traditional statistical methods and simple heuristic rules were used in hospital inventory management. These methods often produced inaccurate forecasts because they did not adequately account for complex factors such as seasonal changes, emergencies, disease outbreaks, and sudden fluctuations in demand (Adedunjoye and Enyejo 2024). Over the past twenty years, the rapid development of machine learning and time series analysis methods has opened up new opportunities in medical demand

Manuscript received: 4 October 2025,
Revised: 11 December 2025,
Accepted: 20 December 2025.

[1]cemozkurt@subu.edu.tr (**Corresponding author**)
[2]23010903094@subu.edu.tr
[3]24010903047@subu.edu.tr
[4]24010903035@subu.edu.tr

forecasting. Algorithms such as tree-based methods (Random Forest, Gradient Boosting, XGBoost, LightGBM), linear regression approaches (Ridge, Lasso), and time series models (ARIMA, Exponential Smoothing) yield good results when used individually (Darshan *et al.* 2025). Academic research and practical applications have shown that a single model does not always deliver optimal performance and may be insufficient in adapting to different data patterns and changing conditions (Punnahitanond *et al.* 2025).

In recent years, hybrid and ensemble model combinations have emerged as innovative solutions to the demand forecasting problem, both in academic settings and industrial applications (Mbonyinshuti *et al.* 2024). Hybrid models combine machine learning and time series methods, leveraging the strengths of each technique, while ensemble approaches systematically combine the forecasts of different models to obtain more stable and reliable results (Vignesh and Vijayalakshmi 2025). Such combination methods capture non-linear relationships that cannot be addressed by individual models, reveal hidden patterns in the data, and significantly reduce prediction errors (Jahin *et al.* 2024). Particularly in the healthcare sector, the use of hybrid and ensemble methods for demand forecasting of hospital supplies has been limited, representing an important area that still needs to be researched (Donkor *et al.* 2024).

The research problem of this study is whether the demand for medical and consumable supplies in hospitals can be predicted more accurately and reliably using hybrid and ensemble model combinations under complex and variable conditions where single prediction models fall short. The main objective of this study is to demonstrate how hybrid and ensemble model combinations can be effectively applied in the demand forecasting of medical and consumable supplies used in hospitals, instead of single models. In this study, historical inventory inflow and outflow data, time series structure, and material usage patterns will be analyzed in detail, and various combinations of tree-based machine learning models, linear regression methods, and time series techniques will be tested. This integrated approach is designed to overcome the limitations of individual models and provide a more reliable, sustainable, and successful forecasting framework for hospital inventory management.

Research Objectives

The main objectives of this research are defined as follows:

- To individually evaluate the success of different machine learning (Random Forest, Gradient Boosting, XGBoost, LightGBM, Ridge, Lasso) and time series models (ARIMA, Exponential Smoothing) in forecasting hospital medical supply demand and to perform a comparative analysis based on performance metrics (MAE, RMSE, MAPE, R^2).
- To design hybrid and ensemble model combinations that combine different machine learning and time series methods in order to overcome the limitations of single model approaches, and to evaluate the prediction success of these combinations.
- Systematically analyze whether the proposed hybrid and ensemble model combinations improve demand forecast accuracy compared to single models, and how they affect model stability and forecast variance.
- To propose a more reliable, economical, and sustainable forecasting framework for inventory management applications in hospitals; to provide recommendations on how this framework can be integrated into practical applications.

Primary Contributions

The main contributions of this study can be summarized as follows:

- Tree-based machine learning models, linear regression methods, and classical time series techniques were comprehensively compared in hospital supply demand forecasting. The performance of each technique was evaluated using standardized metrics.
- Innovative hybrid model combinations integrating machine learning and time series methods have been developed. These hybrid structures overcome the limitations of individual models, providing higher prediction accuracy.
- Various ensemble techniques (bagging, boosting, stacking, weighted averaging) have been systematically applied to optimally combine the predictions of different models. These strategies reduce prediction error and increase system stability.
- Factors specific to the hospital environment, such as seasonal changes, emergencies, and sudden fluctuations in demand, were taken into account in model design. Thus, solutions adapted to the healthcare sector were presented, unlike general industry applications.
- Detailed recommendations regarding model selection criteria, data preprocessing procedures, hyperparameter tuning, and system integration enhance the practical applicability of the study.
- The proposed hybrid and ensemble combinations significantly improve prediction accuracy compared to individual models, thereby contributing to reduced inventory costs and optimized hospital operations.

Structure of the Article

The structure followed in the article begins with the Related Work section, which addresses existing academic studies on demand forecasting techniques, their applications in hospital settings, and the effectiveness of machine learning and ensemble methods. Subsequently, the Materials and Methods section detail the dataset used, data preprocessing steps, and the applied methodology, introducing tree-based models, linear regression methods, time series techniques, and ensemble strategies. In the Hybrid Model Design and Ensemble Strategies sections, combinations of different techniques are systematically created and integrated, and the design principles of each hybrid architecture are compared with the advantages and disadvantages of various ensemble combination methods. In the Results and Discussion section, the performance of all models is compared using standardized metrics (MAE, RMSE, MAPE, etc.), performance improvements of hybrid and ensemble models compared to individual models are presented, detailed comparisons between different model categories are made, and the implications of the findings for hospital operations are outlined by explaining their meaning in the context of the healthcare sector. Finally, the Conclusion section summarizes the main findings of the research, highlights the contributions of the study, provides practical recommendations for hospital managers and operational staff, and outlines guidelines for future research.

RELATED WORK

The healthcare sector faces unique challenges in inventory management and demand forecasting due to its complex structure and critical outcomes. The availability of medical supplies and consumables has a direct impact on the quality and safety of patient care.

Insufficient stock levels can lead to operational disruptions, service interruptions, and even situations that threaten patient safety. Excessive inventory, on the other hand, causes financial problems such as expired shelf life, storage costs, and insufficient capital. In this context, accurately forecasting the demand for medical and consumable supplies in healthcare institutions is of critical strategic importance for cost optimization, operational efficiency, and, most importantly, the provision of uninterrupted, high-quality patient care. In this context, recent literature has been reviewed thematically.

Traditional Time Series Approaches and Their Limitations

Among the classical time series methods commonly used in the literature for forecasting demand for medical and consumable supplies in the healthcare sector are Moving Average (MA), Exponential Smoothing (ES), and Autoregressive Integrated Moving Average (ARIMA) models (Shih and Rajendran 2019; Khalid 2024). While these methods demonstrate a certain degree of success in forecasting future demand based on historical data (Luo *et al.* 2017), they fail to adequately model external factors arising from the dynamic nature of the healthcare sector (Khalid 2024). Particularly in demand series with non-linear relationships and high volatility, the forecasting performance of these models may decline and they may be insufficient in capturing complex patterns (Kolambe 2024; Dachepalli 2025).

In a study predicting hospital admissions during the COVID-19 pandemic, ARIMA and Exponential Smoothing models were used, but limitations in their performance were observed due to the dynamic and non-linear nature of the pandemic (Perone 2021). Similarly, despite the use of models such as ARIMA in forecasting demand in the blood product supply chain, difficulties are encountered in managing the uncertainties arising from the short shelf life and highly variable usage rates of these products (Motamedi *et al.* 2024). Although traditional time series methods have advantages such as usability and low computational cost, especially in short-term forecasting, their inability to fully grasp the complex structure of demand in healthcare services is a significant limitation (Luo *et al.* 2017; Fatima and Rahimi 2024).

These studies demonstrate that classical time series methods have certain advantages in demand forecasting in the healthcare sector; however, they also reveal that nonlinear structures, sudden demand spikes, and material-based heterogeneous usage patterns cannot be adequately captured. In this study, in order to overcome these limitations, a hybrid and ensemble approach integrating machine learning-based models was adopted instead of using classical time series models alone. The goal was to produce more flexible and reliable forecasts for complex demand structures where traditional methods fall short.

The Rise of Machine Learning-Based Approaches

In recent years, machine learning (ML) algorithms have emerged as a powerful alternative for time series forecasting problems and, in particular, demand forecasting in the healthcare sector (Kontopoulou *et al.* 2023). Ensemble learning models such as Random Forest (RF), Gradient Boosting (GBM), XGBoost, and LightGBM stand out for their ability to model complex and non-linear relationships (Qiu *et al.* 2019; Erdebilli and Devrim-Tenba 2022). These models can provide higher prediction accuracy compared to traditional methods and can include a large number of explanatory variables, such as hospital capacity, patient numbers, and seasonal indices, in the analysis (Shern *et al.* 2024).

In predicting peak demand days for cardiovascular diseases, the LightGBM model stood out with a higher AUC value (0.940) compared to other ML models such as logistic regression and SVM (Qiu *et al.* 2019). Similarly, in a study on medical waste prediction, an ensemble consensus regression model using algorithms such as Random Forest, Gradient Boosting, and AdaBoost demonstrated superior performance with a lower RMSE value compared to single models (Erdebilli and Devrim-Tenba 2022). Such tree-based models generally demonstrate strong capabilities in analyzing and predicting complex data structures (Gldoan *et al.* 2023).

Studies using linear models such as Ridge and Lasso also offer limited but valuable contributions, particularly in the context of feature selection and preventing overfitting (Abdul-Rahman *et al.* 2021). The literature emphasizes that machine learning methods are more flexible and adaptable than traditional statistical methods, especially when dealing with complex and high-dimensional data (Fatima and Rahimi 2024). These developments support more accurate decision-making and increased operational efficiency in demand management in the healthcare sector.

Hybrid and Ensemble Models: Potential and Gaps in the Literature

In order to overcome the limitations of single models, hybrid and ensemble modeling approaches have recently gained attention. Combining the strengths of different model families, these approaches generally exhibit superior forecasting performance compared to models used alone (Perone 2021). Ensemble models offer the potential to reduce forecast variance, correct bias, and create more generalizable models by bringing together multiple base learners (Fatima and Rahimi 2024). Such model combinations are particularly promising for demand series in the healthcare sector, which often involve heterogeneous data structures. For example, in a study on energy demand forecasting, hybrid combinations of different time series and machine learning models significantly outperformed individual models.

While much of the existing literature focuses on optimizing the performance of a single model, studies that systematically combine multiple models and comprehensively evaluate their forecasting success are limited (Kontopoulou *et al.* 2023). Particularly in the healthcare sector, there is a need for in-depth analysis of the effectiveness of combinations using different model families in forecasting medical and consumable demand. This points to a significant gap in the literature, as the complex and highly variable demand structures in the healthcare sector suggest that a single model may not always provide the best solution (Qiu *et al.* 2019).

This study aims to fill this gap in the literature by focusing on medical and consumable supply demand forecasting in the healthcare sector. Rather than demonstrating the superiority of a single model, it will compare the forecasting performance of different combinations of classical time series, machine learning, and hybrid approaches. This approach has the potential to contribute to the development of stronger and more consistent forecasting models in healthcare supply chain management, thereby increasing operational efficiency and minimizing costs (Aifuwa *et al.* 2020). Thus, it is assumed that by bringing together complementary information obtained from different model families, more robust and reliable forecasting systems that better adapt to the unique dynamics of the healthcare sector can be created.

Although these machine learning-based studies have made significant progress in capturing non-linear relationships in demand forecasting in the healthcare sector, they mostly focus on individual model performance and address the potential for using different model families together to a limited extent. Unlike the existing literature, this study aims to provide an integrated forecasting framework by systematically combining tree-based methods with linear models, complementing the strengths and weaknesses of individual models.

METHODOLOGY

This section details the methodological framework of the forecasting and decision support system used in hospital inventory management. The primary objective of the study is to generate short-term (three-month) demand forecasts for the future by utilizing historical consumption records on a material-warehouse basis and to convert these forecasts into outputs that can be directly integrated into operational decision-making processes. In this context, all steps, from raw data processing to the creation of monthly time series, from multi-model forecasting approaches to reliability assessment, and from ABC-XYZ segmentation to optimal stock level calculations, are addressed under a comprehensive methodology. The method followed aims to reduce model fragility in healthcare inventories with different consumption regimes, ensure the physical and operational suitability of forecasts, and make the results suitable for internal auditing and reporting (Silva-Aravena *et al.* 2020; Karamshetty *et al.* 2022; Subramanian 2021).

Problem Definition and Notation

The fundamental problem addressed in this study is the accurate and reliable prediction of future short-term consumption behavior for each material based on hospital warehouses. Specifically, the objective is to estimate consumption (output) quantities for the next three months for each material-warehouse combination and to use these estimates to generate decision support outputs for inventory management. The generated forecasts are not merely numerical predictions; they also form the basis for advanced analyses such as ABC-XYZ segmentation, optimal stock level calculation, and automatic identification of problematic materials. This approach directly links the forecasting process to operational policy generation (Balkhi *et al.* 2022; Feibert *et al.* 2019; Polater and Demirdogen 2018).

The system generates monthly consumption time series from raw material movement records and analyzes these time series using an ensemble forecaster. Only physically meaningful movement types are considered in the forecasting process; in this context, transactions in the raw data set are filtered only as G (input) and C (output). Thus, the model operates on a data structure that represents the actual effect of stock movements on inventory levels.

The mathematical notation used throughout this section is defined as follows. Here, m represents the material ID and corresponds to the `HASTANE_MALZEME_ID` field in the application. Similarly, d represents the warehouse ID and is represented by the `AD` field. The time dimension t is defined in terms of the timestamp corresponding to the beginning of the month. The expression $y_{m,d,t}$ used in Equation (1) indicates the output, i.e., the consumption quantity of material m in warehouse d in month t. The forecast horizon is expressed by the parameter h, which in this study covers three-month forward forecasts with $h = 1, 2, 3$.

Data Source and Transformation

The raw dataset used in this study consists of detailed material movement records obtained from a real-world hospital inventory management system through an institutional collaboration. The dataset was provided under a confidentiality agreement. Due to data security and institutional privacy policies, the identity of the data-providing institution and the raw dataset itself cannot be publicly disclosed. The dataset includes fields such as material ID, receipt date, transaction type (inbound or outbound), transaction quantity, and warehouse information.

In order to obtain consistent and reproducible results in forecasting and decision support processes, the raw data was first subjected to a systematic transformation process. In this process, fields unnecessary for analysis were eliminated, numerical variables were converted to appropriate data types, and date fields were converted to timestamp format to ensure temporal consistency (Merkuryeva *et al.* 2019; Subramanian 2021).

To create monthly consumption time series, the exit transactions for each material-warehouse combination were aggregated on a monthly basis. As shown in Equation (1), the consumption quantity for a given month t for a specific material m and warehouse d is defined as the sum of all exit transactions occurring within that month (Mbonyinshuti *et al.* 2022):

$$y_{m,d,t} = \sum_{i \in I(m,d,t)} \text{MIKTAR}_i \qquad (\text{only ISLEM\_TUR} = C) \tag{1}$$

The time series created using the monthly consumption values obtained in Equation (1) were converted to a fixed monthly frequency to make them suitable for the analysis and forecasting process. During this conversion process, the time axis was reindexed in monthly periods for each material-warehouse combination, and it was checked whether there were any outflow transactions in specific months.

This process was carried out as shown in Equation (2). Accordingly, if a consumption observation exists for the relevant month, the time series value is retained as is; if there is no discharge operation in the relevant month, the consumption quantity is assigned a value of zero:

$$y_t \leftarrow \text{asfreq(MS)}, \qquad y_t = \begin{cases} y_t, & \text{if there is consumption} \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

In Equation (2), the expression `asfreq(MS)` represents the resampling of the time series to a monthly start frequency. As a result of this process, the time series is converted into a regular monthly structure, and missing months are explicitly identified. Assigning a value of zero when there is no consumption in the relevant month prevents "no usage" situations from being implicitly lost in the data and allows the model to learn irregular or infrequent usage patterns.

Thanks to this approach, continuous, regular, and comparable monthly time series are obtained for each material-warehouse combination; at the same time, zero consumption periods are preserved as a meaningful signal in statistical and behavioral analyses. Thus, the time series created provide a suitable input structure for both classical time series methods and feature-based machine learning models.

Data Cleaning and Outlier Management

This subsection details how data quality issues that could affect the reliability of forecasting and inventory decisions are addressed. Health inventory data, being derived from operational processes, may contain missing records, incorrect date information, and consumption values that are statistically extreme. Since such problems can lead to misleading patterns and unstable model behavior, especially in time series-based forecasting models, a systematic data cleaning and outlier management process was applied prior to analysis (Karamshetty *et al.* 2022; Gurumurthy *et al.* 2021).

During the preprocessing stage, records with missing information in fields required for analysis were first removed from the dataset. These fields include basic variables such as material ID, warehouse information, transaction date, and transaction amount. Subsequently, quantity fields were converted to appropriate numerical data types, and values with logical inconsistencies were checked. Finally, date fields were parsed into timestamp format, and records with invalid or incorrect date information were eliminated from the dataset, thus ensuring the chronological integrity of the time series.

During the outlier management phase, a top-end trimming (minorize-like) approach was adopted to prevent records with extremely high values, which are rarely observed in consumption amounts, from disproportionately affecting the modeling process. In this approach, the distribution of raw consumption quantities was considered, and the 99.9th percentile value was set as the threshold. As shown in Equation (3), this threshold value is represented by τ, and only observations below this value are retained in the analysis:

$$\tau = Q_{0.999}(x), \qquad x_i^* = \begin{cases} x_i, & x_i \leq \tau \\ \text{remove}, & x_i > \tau \end{cases} \tag{3}$$

In Equation (3), x_i represents the raw consumption (transaction amount) value, while $Q_{0.999}(x)$ represents the upper 99.9th percentile of the entire consumption distribution. According to this definition, observations satisfying the condition $x_i \leq \tau$ are retained in the data set, while extreme outliers above the threshold value are excluded from the analysis. This method aims to prevent the model from developing excessive sensitivity without completely suppressing the effect of outliers, while preserving the overall structure of the distribution.

This outlier cleaning strategy offers a balanced solution that maintains both statistical robustness and operational realism, particularly in healthcare inventory data where high-volume outputs are rarely used but occasionally observed. Thus, the resulting time series achieve a more stable and reliable input structure for both classical statistical analyses and machine learning-based forecasting models (Ingle *et al.* 2021; Kim *et al.* 2023).

Material-Based Statistical and Behavioral Feature Extraction

This subsection explains how statistical and behavioral features that quantitatively summarize the past consumption behavior of each material are extracted. The aim is to improve prediction performance and generalizability by including attributes that reflect the general usage character of the material in the model, rather than using only the past values of the time series. This approach enables the differentiation of materials that have the same lagged values but exhibit different usage patterns (Kim *et al.* 2023; Subramanian 2021).

For each material–storage combination, basic descriptive statistics such as total number of entries and exits, average consumption amount, maximum and minimum values, and standard deviation were calculated. However, the coefficient of variation (CV) was used to evaluate consumption behavior independently of scale. As shown in Equation (4), CV is defined as the ratio of the standard deviation to the mean and measures the relative variability in consumption:

$$CV = \frac{\sigma(y)}{\mu(y)} \qquad (\mu(y) > 0) \tag{4}$$

In Equation (4), σ represents the standard deviation of monthly output quantities, while μ represents the mean value of the same series. The CV metric prevents absolute variance from being misleading in materials with low averages, enabling a comparative assessment of consumption stability.

Seasonality strength is defined to quantitatively express the degree of seasonal fluctuation in consumption throughout the year. In this context, the seasonality indicator is obtained by normalizing the variability of average consumption values by month. As shown in Equation (5), this metric is calculated based on the ratio of the variation between monthly averages to the overall average:

$$S = \frac{\sigma(\bar{y}_{ay})}{\mu(\bar{y}_{ay})} \qquad (\mu > 0) \tag{5}$$

Here, $\bar{y}_{ay}$ represents the average consumption calculated for each month, while $\bar{y}$ represents the overall average obtained throughout the entire period. Higher S values indicate strong seasonal fluctuations and show that the forecasting process is relatively more complex.

The trend slope was obtained by applying linear regression to the monthly total consumption values in order to determine the general upward or downward direction of the consumption series over time. As shown in Equation (6), the slope of the consumption values against the time variable is used as a trend indicator:

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t \tag{6}$$

In this equation, the coefficient β_1 represents the trend slope, indicating whether consumption shows an increasing or decreasing trend over time.

Finally, the regularity score is defined to measure the degree to which a material is used regularly. As shown in Equation (7), regularity is calculated as the ratio of the months in which consumption occurred to the total number of months:

$$R = \frac{\text{Number of months consumption was observed}}{\text{Total number of months}} \tag{7}$$

This metric plays a critical role in both prediction reliability and inventory policy design by enabling the differentiation of irregular and infrequently used materials.

Time Series Feature Engineering

This subsection details the feature engineering process applied to transform monthly consumption time series into an input space suitable for supervised learning models. In time series forecasting problems, the direct use of past observations is often insufficient; instead, derived features representing temporal dependencies, short-term trends, and calendar effects must be included in the model. Within the scope of this study, a comprehensive feature vector consisting of lagged values, rolling statistics, short window trends, and calendar variables was created for each month (Kim *et al.* 2023; Ingle *et al.* 2021).

The feature vector created for each time step primarily includes calendar components. In this context, the variables "month(t)", "year(t)", and "quarter(t)" are used to enable the model to learn monthly, yearly, and quarterly seasonal effects. Additionally, the variable k, an increasing index starting from the beginning of the time series, is defined to represent long-term trends through a linear time indicator.

To capture temporal dependencies, lagged values of past consumption for each month are added to the feature vector. In this study, consumption amounts from one-, two-, and three-months prior were defined as y_{t-1}, y_{t-2}, and y_{t-3}, respectively, and the short-term autocorrelation structure was targeted for transfer to the model.

Rolling statistics were calculated to represent the short-term consumption level and volatility. As shown in Equation (8), the three-month rolling average $RM_3(t)$ is defined as the arithmetic mean of the previous three observations, including the relevant month:

$$RM_3(t) = \frac{1}{3}\sum_{i=0}^{2} y_{t-i} \tag{8}$$

In Equation (8), y_{t-i} represents the consumption amount i months prior to time t. This metric represents the short-term consumption level in a smoothed form, reducing the impact of sudden spikes.

Additionally, a three-month rolling standard deviation was calculated to measure short-term consumption volatility. As shown in Equation (9), the $RS_3(t)$ value is defined as the square root of the average of the squares of deviations around the rolling average:

$$RS_3(t) = \sqrt{\frac{1}{3}\sum_{i=0}^{2}\left(y_{t-i} - RM_3(t)\right)^2} \tag{9}$$

Equation (9) quantitatively expresses the magnitude of short-term consumption fluctuations and enables the model to distinguish between stable and volatile periods.

The short window trend is defined to capture the direction of increase or decrease in the recent past. In this context, the slope value obtained by applying a linear regression on the last three observations is expressed as $\text{Trend}_3(t)$ as shown in Equation (10):

$$\text{Trend}_3(t) = \text{slope}(y_t, y_{t-1}, y_{t-2}) \tag{10}$$

This slope value represents the recent consumption direction independently of long-term trends and ensures that short-term changes are incorporated into the forecasting process.

All features defined in this subsection are explicitly generated within the system and are perfectly aligned with the feature space used during the training phase. The same feature generation scheme is maintained for forecasts covering the next three months; feature vectors for future periods are created using lagged values derived from the latest observations, rolling statistics, and trend components. This ensures structural consistency between the training and forecasting phases, enhancing the model's stability and generalizability.

Multi-Model Forecasting and Ensemble Combination

This subsection details how monthly consumption time series are handled within a multi-model forecasting framework and how models with different learning biases are combined to obtain more stable forecasts. Health inventory consumption data can exhibit heterogeneous behaviors such as high variation, irregular usage, sudden spikes, and regime shifts, making a forecasting approach based on a single model family often insufficient. Therefore, the study adopts an ensemble structure that combines linear, tree-based, and boosting-based models (Chien *et al.* 2023; Sina *et al.* 2023).

The system uses XGBoost and LightGBM models when appropriate libraries are available in the working environment; if these libraries are not available, it constructs an equivalent ensemble space using Random Forest and Gradient Boosting algorithms representing the same model class. Additionally, linear-regularized models such as Ridge and Lasso are included in the ensemble to capture linear components in consumption behavior. This approach aims to achieve more balanced prediction performance across different consumption regimes through model diversity.

Converting Time Series to Regression (Supervised TS Formulation): This subsection explains the process of converting monthly consumption time series into a structure that can be processed by supervised learning algorithms. In time series forecasting problems, it is known that each observation is not solely based on past values; it must also be considered alongside calendar effects, short-term statistics, and local trend information. Therefore, in this study, the time series problem is transferred to a regression framework by defining a comprehensive feature vector for each time step.

For the monthly consumption series y_t, the feature vector created at each time step t is defined as shown in Equation (11):

$$x_t = [\underbrace{\text{month}(t), \text{year}(t), \text{quarter}(t), k}_{\text{calendar + index}}, \underbrace{y_{t-1}, y_{t-2}, y_{t-3}}_{\text{delays}}, \underbrace{\mu_t(3), \sigma_t(3)}_{\text{rolling}}, \underbrace{s_t(3)}_{\text{short trend}}] \tag{11}$$

The variables $\text{month}(t)$, $\text{year}(t)$, and $\text{quarter}(t)$ in Equation (11) ensure that the monthly, annual, and quarterly seasonal effects on consumption are incorporated into the model. The k variable represents the increasing time index from the beginning of the time series and allows long-term trends to be learned through a linear time indicator. The lagged variables y_{t-1}, y_{t-2}, y_{t-3} reflect the short-term autocorrelation structure, ensuring that past consumption information is directly included in the model.

Short-term statistical properties are calculated based on the last three observations. The three-month moving average $\mu_t(3)$ and moving standard deviation $\sigma_t(3)$ are defined as shown in Equation (12):

$$\mu_t(3) = \frac{1}{3}\sum_{i=0}^{2} y_{t-i}, \qquad \sigma_t(3) = \sqrt{\frac{1}{3}\sum_{i=0}^{2}\left(y_{t-i} - \mu_t(3)\right)^2} \tag{12}$$

These metrics represent the short-term level and volatility of consumption, ensuring that sudden spikes and temporary fluctuations are incorporated into the model in a balanced manner.

The short-window trend is defined by the slope value obtained by applying a linear fit to the last three observations. This process is performed as shown in Equation (13):

$$y_{t-i} \approx ai + b \quad (i = 0, 1, 2), \qquad s_t(3) = a \tag{13}$$

Equation (13), where a represents the trend coefficient reflecting the short-term consumption trend. This value ensures that the recent upward or downward direction is incorporated into the model independently of long-term trends.

In practice, this structure was implemented through the variables `lag_1-3`, `rolling_mean_3`, `rolling_std_3`, and `trend_3`. In addition, attributes such as the coefficient of variation (CV) representing material behavior, seasonality score, long-term trend slope, and regularity were also added numerically to the feature vector. In this way, the consumption context of different materials with the same lag values is clearly presented to the model, enabling the model to learn material-specific behaviors (Kim *et al.* 2023).

Mathematical Foundations of the Models Used

This subsection presents the mathematical foundations of the prediction models used in the study. The models were selected to have different learning biases and positioned within the ensemble structure to complement each other.

In linear regression models, the objective is to estimate the coefficients representing the linear relationship between the input features and the target variable. This fundamental relationship can be expressed as shown in Equation (14):

$$\hat{y} = X^T \beta \tag{14}$$

In this model, β represents the coefficient vector to be estimated. In high-dimensional and correlated feature spaces, the classical least squares approach can lead to overfitting. Ridge and Lasso regularizations have been used to mitigate this problem.

Ridge regression adds a penalty term based on the L_2 norm to limit the magnitude of the coefficients. As shown in Equation (15):

$$\min_{\beta} \sum_{i=1}^{n} (y_i - x_i^T \beta)^2 + \lambda \|\beta\|_2^2 \tag{15}$$

This approach reduces the model's variance by bringing the coefficients closer to zero, thereby producing more stable estimates.

Lasso regression, on the other hand, produces sparse solutions using a penalty term based on the L_1 norm. This indirectly provides a feature selection mechanism by setting the coefficients of unimportant features to zero. Equation (16) illustrates this structure:

$$\min_{\beta} \sum_{i=1}^{n} (y_i - x_i^T \beta)^2 + \lambda \|\beta\|_1 \tag{16}$$

Ridge and Lasso models form a strong baseline, particularly in feature spaces containing partially linear relationships such as calendar variables, lagged consumption values, and rolling statistics. Using these models in conjunction with tree-based methods allows the balanced capture of linear and nonlinear components in the ensemble structure (Kim *et al.* 2023).

Random Forest, one of the tree-based methods used in this study, has been included in the ensemble structure to obtain stable estimates, especially for high-variance and noisy consumption time series. Since health inventory consumption data can contain sudden usage spikes and irregular usage periods, models based on a single decision tree often face high variance problems. The Random Forest approach aims to mitigate this limitation through bootstrap sampling and averaging mechanisms (Mbonyinshuti *et al.* 2022; Kim *et al.* 2023).

The Random Forest model produces the final output by averaging the predictions of numerous decision trees trained on different subsets of data created using the bootstrap method. This structure is expressed as follows, as shown in Equation (17):

$$\hat{y}_{RF}(x) = \frac{1}{T} \sum_{t=1}^{T} f_t(x) \tag{17}$$

Equation (17) shows that $f_t(x)$ represents the prediction generated by the t-th decision tree for the input feature vector x, while T denotes the total number of trees in the ensemble. Each decision tree is trained on a different subset of the original dataset, sampled using the bootstrap method; additionally, randomly selected feature subsets are used in the splitting operations at tree nodes. These two randomness mechanisms significantly reduce the variance of the ensemble average by lowering the correlation between trees.

While a single decision tree typically exhibits low bias but high variance, the Random Forest ensemble suppresses this high variance through averaging. This feature enables more stable and reliable predictions, especially in health inventory time series where consumption amounts fluctuate irregularly, very high values are rarely observed, and the noise level is high. Therefore, the Random Forest model plays a critical role as a variance-reducing component within the ensemble structure.

Gradient Boosting, another tree-based method used in this study, was included in the ensemble structure to gradually correct systematic errors that arise in consumption estimation. While the Random Forest approach focuses on reducing variance, the Gradient Boosting model primarily aims to reduce bias and gradually learns the error components that previous models could not explain. This feature provides a significant advantage, especially in complex consumption patterns where trends, seasonality, and lagged interactions are observed together (Sina *et al.* 2023; Kim *et al.* 2023).

The Gradient Boosting approach builds the model incrementally. In the first step, an initial model that minimizes the loss between observed values and predictions is defined. This process is expressed as shown in Equation (18):

$$L(y_i, c) \tag{18}$$

In Equation (18), $L(y_i, c)$ represents the loss between the actual consumption value y_i and the constant estimate c. Following the initial model, the model is updated incrementally. At each step, a new weak learner is added that attempts to explain the residuals (negative gradients) of the previous model. This update process is shown in Equation (19):

$$F_m(x) = F_{m-1}(x) + \nu h_m(x) \tag{19}$$

Here, $h_m(x)$ represents the new weak learner attempting to explain the error component that the previous model could not predict, while ν represents the learning rate. The learning rate reduces the risk of overfitting by controlling the extent to which each new model contributes to the ensemble output. Thanks to this stepwise structure, the Gradient Boosting model gradually captures components that could not be explained in previous steps and systematically improves prediction accuracy.

Due to these characteristics, Gradient Boosting plays an important role as a bias-reducing component within the ensemble structure.

XGBoost and LightGBM are enhanced versions of the Gradient Boosting approach in terms of regularization and computational efficiency. These models are designed to offer stronger generalization performance, particularly in high-dimensional feature spaces and complex interaction structures. In this study, these models were included in the ensemble structure to more effectively capture nonlinear and interactive consumption dynamics (Cao and Gui 2019; Kim *et al.* 2023).

The optimization process in XGBoost and LightGBM is based on jointly minimizing the loss function, which measures data fit, and a regularization term that penalizes model complexity. This objective function is generally defined as shown in Equation (20):

$$L = \sum_{i} L(y_i, \hat{y}_i) + \sum_{k} \Omega(f_k) \tag{20}$$

In Equation (20), $L(y_i, \hat{y}_i)$ represents the loss between the actual consumption value and the model estimate, while the term $\Omega(f_k)$ represents the model complexity of the k-th tree. A typical regularization function used for tree complexity is given in Equation (21):

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \tag{21}$$

In this expression, T represents the number of leaves in the tree, w represents the weights associated with leaf nodes, and γ and λ represent the regularization coefficients that penalize model complexity. These penalties imposed on the number of leaves and weights prevent the model from learning overly complex structures, thereby improving its generalization ability.

Thanks to this regularization mechanism, XGBoost and LightGBM can effectively model complex consumption dynamics such as interactions between consumption values, seasonal effects, and irregular usage patterns. On the code side, if these libraries are not available in the working environment, the model is preserved by substituting Random Forest or classic Gradient Boosting algorithms representing the same model class.

Time Series Cross-Validation (TS-CV) and Model Selection with MAE

This subsection explains how the performance of prediction models is evaluated in a manner appropriate to the nature of time series. In time series problems, classical shuffled cross-validation approaches can cause information about the future to influence past model parameters, a situation referred to in the literature as "data leakage". Since such leakage can cause model performance to appear more optimistic than it actually is, this study adopts a time-axis-sensitive validation strategy (Shaub 2020; Kim *et al.* 2023).

In this context, the system gradually expands the training window along the time axis using the TimeSeriesSplit approach and positions the validation set chronologically ahead of the training set at each fold. Thus, observations after time t are not used in learning the model's parameters at time t, and the forward prediction scenario is simulated realistically.

Mean Absolute Error (MAE) was chosen as the error metric for evaluating model performance. Equation (22) provides the mathematical definition of MAE:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{22}$$

In Equation (22), y_i represents the observed actual consumption value, while $\hat{y}_i$ represents the estimate generated by the model. The MAE metric provides a more robust measure than MSE against singular jumps in time series that may contain extreme values, such as stock consumption, because it does not involve squaring. This feature makes MAE a suitable performance criterion, especially for healthcare inventory data where both high-volume consumption and zero consumption can occur.

Selecting the Top 3 Models and Weighted Ensemble

As a result of the time series cross-validation process, MAE-based performance scores were obtained for each candidate model. Using these scores, the models with the lowest error values were selected and used to create the ensemble structure. Specifically, as shown in Equation (23), the first three models with the smallest MAE values were determined:

$$\mathcal{B} = \text{Top3}\left(\arg\min_{j} MAE_j\right) \tag{23}$$

In Equation (23), $\mathcal{B}$ represents the set of models that will be included in the ensemble, while MAE_j denotes the error score obtained from the cross-validation of the j-th model.

The predictions of the selected models are combined using a performance-based weighting approach. The weights are defined as inversely proportional to the error, as shown in Equation (24) (Pawlikowski and Chorowska 2020; Shaub 2020):

$$w_j = \frac{1}{MAE_j + \varepsilon} \tag{24}$$

Here, ε represents a small constant value added to ensure numerical stability and prevent division by zero. The obtained weights are normalized, and the final ensemble estimate is calculated as shown in Equation (25):

$$\hat{y}(x) = \frac{\sum_{j\in\mathcal{B}} w_j \hat{y}_j(x)}{\sum_{j\in\mathcal{B}} w_j} \tag{25}$$

The statistical interpretation of this approach is that models with low MAE values have smaller overall error levels and are therefore targeted to reduce the expected absolute error by receiving higher weights in the combination process.

Constraining Negative Predictions (Physical Feasibility)

Since consumption quantities cannot take negative values by nature, a one-sided constraint has been applied to all model predictions to preserve the physical meaningfulness of the prediction outputs. This constraint ensures that predicted values below zero are clipped to zero, as shown in Equation (26):

$$\hat{y} \leftarrow \max(\hat{y}, 0) \tag{26}$$

This process can be interpreted as a non-negativity constraint applied in the output space and is applied both in individual model forecasts and in the ensemble output vector. This ensures that stock and consumption forecasts are consistent with physical reality.

Multi-Step ($H = 3$) Forecast Generation

In this study, the forecast horizon is set to $H = 3$ months. To generate forecasts for future periods, feature vectors for future periods are constructed using lagged values derived from the latest observations, rolling statistics, and trend components. These feature vectors are defined as $x_{t+1}, x_{t+2}, x_{t+3}$, respectively (Cao and Gui 2019).

The selected ensemble model directly produces multi-step forecasts based on these feature vectors. This process can be expressed as shown in Equation (27):

$$[\hat{y}_{t+1}, \hat{y}_{t+2}, \hat{y}_{t+3}] = \text{Ensemble}(x_{t+1}, x_{t+2}, x_{t+3}) \tag{27}$$

The code creates a `future_df` for the next three months and aligns it with the training feature space using `X_test = future_df[X_train.columns]`.

Robust Fallback Strategies

Health inventory consumption series can exhibit heterogeneous behaviors such as sparse usage, sudden spikes, institution/clinic-driven regimen changes, and seasonal fluctuations. Therefore, relying on a single model family increases the risk of "single-model fragility" (the model becoming fragile under certain data regimes). In this study, the prediction engine is designed with a multi-layered fallback architecture (Sina *et al.* 2023; Kim *et al.* 2023).

Primary Layer: Ensemble Learning: When the length of the time series created at monthly frequency is $T \geq 6$ months, the system primarily performs prediction through a feature-based ensemble learning architecture. At this stage, the time series is modeled not directly from raw values but through an explanatory feature space representing the series' temporal structure and short-term dynamics. The feature vector created for each time step includes calendar information (month, year, and quarter), the time index, lagged variables representing past consumption (values from the last one, two, and three months), local statistics such as the rolling average and standard deviation for the last three months, and the local trend slope calculated over a short window. Additionally, statistical measures summarizing the historical usage behavior of the relevant material, such as the coefficient of variation (CV), regularity of usage, and seasonality indicator, are also included in this feature space as numerical attributes. This ensures the model is sensitive not only to the latest values in the time series but also to the material's long-term behavioral characteristics.

Multiple learners are trained in parallel on this expanded feature space. Gradient-boosted tree-based methods such as XGBoost and LightGBM are used in the application environment whenever possible; when these libraries are not available, Random Forest and Gradient Boosting algorithms, which represent the same model family, are used as backups. The fundamental purpose of this model diversity is to capture nonlinear interactions, sudden jumps, or smoother trend structures that may arise in different consumption regimes using models with different biases.

Model selection and performance evaluation are performed using a TimeSeriesSplit-based cross-validation approach to prevent information leakage in time series. In this method, the training window is gradually expanded along the time axis, and each validation step is performed using only historical data. The Mean Absolute Error (MAE) is preferred as the error metric, and the most successful models are determined based on the average validation error obtained for each model. The predictions of the models with the lowest error are combined using weights inversely proportional to the corresponding error values, thus yielding the final ensemble prediction. This weighted combination strategy aims to reduce the impact of systematic errors that a single model may make under certain data regimes and to increase overall generalization capability by allowing more reliable models to contribute more to the prediction.

Secondary Layer: Exponential Smoothing Backup: The ensemble learning layer may fail in practice due to unexpected numerical or structural errors during feature extraction, model training, or prediction generation. In such cases, a complete halt in the prediction process is unacceptable for operational decision support systems. Therefore, the system incorporates the Holt–Winters Exponential Smoothing approach from classical time series methods as a backup mechanism in the second layer. This method tends to produce stable and interpretable results, especially for medium and long-length series, thanks to its fewer parameters and its ability to directly model the level, trend, and, if necessary, seasonality of the time series.

Whether or not the seasonality component is included in the model is determined conditionally. If the time series is of sufficient length (at least $T \geq 24$ months in practice) and the seasonality indicator, which shows a meaningful seasonal fluctuation in the historical behavior of the series, is above a certain threshold value, the seasonal component is added to the model. Otherwise, a simpler structure working only with level and trend components is preferred. The main objective of this approach is to ensure the continuity of the system's forecasting by compensating for the breaks that may occur in machine learning-based methods when the data is of medium or long length with a more classical and stable time series model.

Third Layer: Simple Scaled Average / Last-Value Heuristic: When the time series length is very short, i.e., especially in cases where $T < 6$ months, the risk of both machine learning-based methods and parametric time series models overfitting or producing unreasonably volatile predictions increases significantly. In such data-starved scenarios, the system consciously adopts the principle of "safe heuristics over complex modeling." If at least a few observations are available for the relevant material, the last observed consumption value is scaled by a specific coefficient (approximately 50% in practice) to produce a fixed projection. If the time series is effectively empty or there is no meaningful historical data, the estimates are fixed to zero, and these outputs are also reported with a very low confidence label.

While not numerically perfect, this approach ensures that the system exhibits a more cautious operational behavior by preventing it from producing random or overly confident but unfounded estimates in the absence of data.

When considered together, this three-layered structure presents a robust prediction architecture that enables the system to behave in an adaptable manner to data length and data quality without being dependent on a single model family. Thus, the goal is for the prediction pipeline to operate seamlessly and reliably in healthcare inventories with heterogeneous consumption patterns.

Prediction Confidence Score:

Simply producing point estimates is not sufficient for prediction outputs to be used in real-world inventory management and supply planning processes. It is also necessary to quantitatively express the conditions under which these predictions are made and the degree to which they are based on reliable information. Especially in critical areas with high operational risk, such as healthcare inventory, ignoring forecast uncertainty can lead directly to costly outcomes such as stock-outs or excess inventory. Therefore, this study defines a rule-based and explainable forecast confidence score accompanying the forecasts generated for each material.

This score is not a machine learning output; it is designed as an interpretable quality measure obtained by jointly evaluating a set of statistical indicators that reflect the amount of data, behavioral stability, timeliness, and seasonal complexity. Thus, decision-makers can obtain a quantitative answer not only to the question "How much will we consume?" but also to the question "How much can I trust this forecast?"

Components and Normalization: In order to make the different indicators calculated for each material comparable, the system first reduces all components to the range $[0, 1]$ and then calculates the average of these values to produce the overall confidence score. This approach allows criteria with different scales and meanings to be combined under a single unified quality metric.

The first component of the confidence score is the data quantity factor, which represents the amount of data on which the prediction is based. This factor is defined in Equation (28) below, using T to denote the length of the time series:

$$f_{data} = \min\left(\frac{T}{24}, 1\right) \tag{28}$$

In this formula, T represents the number of monthly observations available for the relevant material. The value of 24 used in the denominator was chosen as an intuitive reference, assuming that at least two years of observation history is a reasonable “sufficiency threshold” to capture annual seasonal cycles. If the time series length is 24 months or longer, this factor takes the value 1, and the data quantity is considered sufficient. Conversely, for shorter series, this ratio remains below 1, and the confidence score is suppressed downward due to data scarcity.

The second component represents regularity of use. This measure is derived from the “proportion of months with zero consumption” in the monthly consumption series for the relevant material and is used directly as a normalized value between 0 and 1. A high regularity value indicates that the material is consumed regularly in most months and therefore exhibits predictable behavior. Conversely, low regularity indicates that the series has infrequent and irregular usage and that prediction uncertainty should naturally be higher.

The third component is a measure representing the stability of variation in the series and is based on the coefficient of variation (CV). CV is a dimensionless statistic defined as the ratio of the standard deviation to the mean, measuring the relative volatility of the series. In this study, CV was converted into a confidence score using a three-interval interval scoring method to enhance practical interpretability, rather than being used directly as a continuous function. If the CV value is below 0.5, the series is considered relatively stable, and in this case, a score of 0.8 is assigned to this component. If the CV is between 0.5 and 1.0, it is assumed to indicate moderate volatility, and the component score is taken as 0.6. In high volatility regimes where the CV is 1.0 or above, the series is considered highly irregular, and this factor contributes only a lower score of 0.3. This threshold directly reflects the assumption that point estimates are naturally less reliable in high-variance series.

The fourth component represents the recency of the data. The magnitude Δ used here indicates the number of days between the last output process for the relevant material and the current date. If a material has not been used for a very long time, predictions based on its past behavior are less likely to reflect current operational reality. Therefore, materials with a small Δ value, i.e., those used recently, receive a higher score. In practice, this factor is taken as 1.0 for $\Delta \leq 90$ days, reduced to 0.7 for values between 90 and 365 days, to 0.4 for values between 365 and 730 days, and to 0.1 for materials that have not been used for more than two years. This gradual decrease directly integrates the concept of “temporal decay” of information into the confidence score.

The fifth and final component is the indicator representing seasonal complexity, denoted by S. This magnitude is a normalized measure summarizing the relative intensity of seasonal fluctuations in the series. In series with a weak seasonal effect, i.e., $S < 0.3$, this factor takes a high value such as 0.8; in series with moderate seasonality ($0.3 \leq S < 0.6$), it drops to 0.6; and in series with strong seasonal fluctuations ($S \geq 0.6$), it is taken as 0.4. The basic assumption here is that as seasonality increases, the behavior of the series becomes more complex and, consequently, the uncertainty of short-term forecasts rises.

Overall Score and Verb Levels: All of the components defined above are combined to calculate a single composite confidence score for each material. The overall confidence score C is defined as the arithmetic mean of the total K components in Equation (29):

$$C = \frac{1}{K} \sum_{k=1}^{K} f_k \tag{29}$$

This expression defines f_k as the normalized sub-scores corresponding to data quantity, regularity, variation stability, timeliness, and seasonal complexity, respectively, and $K = 5$ was used in this study. The reason for choosing the arithmetic mean is to prevent any single factor from overly dominating the score by giving equal weight to all components and to obtain a more balanced reliability assessment.

This calculated continuous value is then converted into four ordinal confidence levels (e.g., “very high”, “high”, “medium”, “low”, “very low”) for easier interpretation by decision-makers. This conversion ensures that the system outputs are directly usable not only numerically but also in managerial and operational reporting contexts (Subramanian 2021).

ABC–XYZ Segmentation

The outputs of the forecasting module are not limited to producing total future consumption quantities; they also reveal a rich set of information reflecting the extent to which the demand behavior of each material is predictable. From an operational inventory management perspective, the question “how much of which material will be consumed?” is not sufficient; it is also necessary to answer the questions “which materials have more stable demand, and which are more volatile and uncertain?” In order to address these two dimensions together in this study, the ABC classification, commonly used in classical inventory literature, was combined with the XYZ classification, which is based on demand predictability, to create a combined ABC–XYZ segmentation for each material. The application aims to ensure that the results obtained are transparent and reproducible in internal auditing, reporting, and decision support processes, thanks to the “rule-based” and deterministic definition of classification thresholds.

ABC: For each material in the ABC dimension, the sum of the estimated consumption quantities for the next three months is used as a proxy variable representing the relative “value” or “intensity” of the relevant item in the inventory system. This total value, V_m, is obtained by summing the three-month forecasts generated for material m, and all materials are ranked from largest to smallest according to this magnitude. Subsequently, the cumulative share of each material in the total forecast is calculated, and the classification is performed based on the position of this share within the total.

In this approach, materials in the top group with a cumulative share up to 80% of the total are labeled as Class A because they represent a large portion of the total consumption volume. Materials with a cumulative share between 80% and 95% form the Class B group, which has a medium consumption profile. Materials in the bottom 5% of the total, with a relatively low consumption volume, are classified as Class C. The selection of these thresholds is consistent with the classical ABC approach, which is based on the Pareto principle, providing a prioritization logic that directs resources and attention primarily to Class A items in inventory management.

XYZ: The XYZ dimension aims to measure the stability and predictability of demand behavior rather than the consumption quantity of materials. To this end, the coefficient of variation (CV) calculated for each material and the regularity indicator representing usage regularity are used together. CV measures the relative volatility of the series, while regularity reflects the degree of continuity in consumption over months. Evaluating these two metrics together provides a more balanced classification that accounts not only for the magnitude of fluctuations but also for the structural continuity of the series.

Within this framework, materials with a CV value below 0.5 and a regularity value above 0.7 are classified as Class X, as they exhibit both low volatility and high usage continuity. This group represents the most predictable demand and the most reliable items for inventory planning. Materials with a CV value below 1.0 but a regularity value above 0.4 are labeled as Class Y, assuming they have medium volatility and partial regularity. This group represents a transition zone with medium uncertainty in terms of predictability. Materials that do not meet these conditions, i.e., those with irregular and complex behavior due to high variation or low regularity, are classified as Class Z and constitute the most problematic group in terms of demand forecasting.

Calculating the Optimal Stock Level

ABC–XYZ segmentation is used in this study not only to classify materials but also to systematically determine the stock management policy to be applied for each segment. In other words, the segment label obtained is treated directly as a "decision parameter," and different safety stock ratios, maximum stock levels, and control frequencies are defined for each segment. This approach ensures the establishment of an adaptable and risk-sensitive structure that takes into account both consumption volume and the predictability of demand behavior, rather than applying a uniform policy in inventory management. In practice, the matching between the segment and the inventory policy is defined through a clear and deterministic dictionary structure, ensuring that the results produced by the system are verifiable and easily integrated into corporate procedures (Goncalves *et al.* 2020).

Demand Representation: The base demand level used in inventory planning calculations for each material is obtained by taking the average of three-month forecasts representing the short-term forecast horizon. This approach aims to mitigate the impact of short-term volatility and produce a more balanced representation of "typical monthly demand" by preventing excessive reliance on a single month's forecast value. Thus, both sudden spikes and temporary declines do not disproportionately affect inventory level decisions; instead, calculations are based on a more stable reference level.

Safety Stock, Maximum Stock and Reorder Point: The safety stock and maximum stock level to be determined for each material are calculated based on policy coefficients defined according to the ABC–XYZ segment to which it directly belongs. In this context, the segment label represents a kind of "risk profile"; for example, higher safety margins are anticipated for segments that are both high-volume and low-predictability, while more limited buffer levels are considered sufficient for low-volume and more stable segments.

The safety stock can be interpreted as a buffer level obtained by applying a segment-based safety factor to the typical monthly demand for the relevant material, while the maximum stock level represents the upper limit that the stock is allowed to reach, obtained by scaling the same reference demand size with a wider multiplier.

The reorder point is defined in the system based on a rule that is practical and operationally easy to implement. In this study, the reorder threshold is considered as a fixed percentage of the maximum stock level, and this percentage is directly applied as a safety percent in practice. This choice aims to balance both the risk of excessive stock accumulation and the risk of stock depletion by automatically triggering an order when the stock level falls below a certain safe margin.

Seasonality Adjustment: The consumption behavior of some materials involves significant seasonal fluctuations. In such cases, stock levels determined by fixed coefficients may be insufficient to meet seasonal demand peaks. Therefore, for materials with a high seasonal effect, the safety stock and maximum stock levels are adjusted upward by an additional multiplier. This adjustment aims to reduce the risk of stock depletion, particularly in clinical usage scenarios where sudden and intense demand spikes occur during specific periods. Thus, the system indirectly reflects not only the average demand level but also the structural fluctuations in the temporal distribution of demand in its stock policies.

Problem Material Identification and Intervention Strategies

A critical step in operational decision support is to automatically flag risky or problematic materials on the same line where forecasts are generated. The system labels a material as "problematic" when any of the following conditions are met **?**:

- Low or very low reliability level
- Forecast quantity is zero
- $CV > 1.5$ (highly variable)
- Last use > 365 days (long period of non-use)

Behavioral Problem Groups: Problem materials are divided into subgroups based on their typical behaviors to facilitate intervention design:

- **OLD_USE**: Materials that have not been used for a very long time.
- **HIGH_VARIATION**: Materials exhibiting very high consumption volatility.
- **IRREGULAR_USE**: Materials with low regularity of usage.
- **INSUFFICIENT_DATA**: Materials with very few historical outputs.
- **COMPLEX_PATTERN**: Materials that do not fit into the above categories and exhibit complex or mixed behavior.

Action Recommendations and Prioritization: This subsection explains how actionable recommendations and priority levels are generated for each problem group, going beyond the identification of problematic materials. The goal is to transform forecasting and classification outputs from merely descriptive reports into a decision support layer that directly supports operational decision-making processes. This approach elevates the system from a "passive reporting" level to an "active and directive decision support" level.

The system determines the problem type by jointly evaluating the forecast results, uncertainty indicators, behavioral characteristics (variation, regularity, trend), and inventory classifications (ABC–XYZ) obtained for each material–warehouse combination. As a result of this comprehensive evaluation, predefined but data-triggered action templates are activated for each problem type.

For example, for materials that exhibit long-term low or zero consumption and whose predicted future consumption is also negligible, removal from stock or evaluation of alternative uses is recommended. Such materials, especially if they are in the C or Z class, can negatively impact warehouse space efficiency and inventory carrying costs. Therefore, the system flags this group as a low operational priority but strategically important problem requiring cleanup.

In contrast, for materials exhibiting high variation and irregular consumption behavior, action recommendations are generated to increase safety stock (buffer levels) or re-adjust reorder points. In such cases, the system proposes a more protective policy aimed at minimizing stock-out risk, taking into account prediction error and uncertainty levels.

When data length is insufficient or consumption patterns cannot be modeled statistically reliably, the system flags these materials as "high uncertainty" and adopts an approach that prioritizes expert evaluation over automatic decision-making. The recommended action plan for such materials is to initiate a manual review process alongside temporary conservative stock policies.

Each generated action recommendation is labeled with a priority score, taking into account the operational impact of the relevant problem, the expected risk level, and the potential cost outcome. This prioritization enables managers to focus limited resources on the most critical materials and systematizes the decision-making process. Thus, the outputs obtained from the forecasting module are transformed into actionable, traceable, and justifiable decision recommendations.

RESULTS AND DISCUSSION

Medical supply demand forecasting in healthcare institutions is a decision problem characterized by high uncertainty, dynamism, and multiple factors. Sudden changes in patient numbers, emergency department workloads, epidemic periods, updates to clinical protocols, and administrative decisions directly and often unpredictably affect consumption patterns. Therefore, healthcare inventory data are often described in the literature as irregular, sparse, highly variable, and prone to anomalies (Merkuryeva *et al.* 2019; Subramanian 2021; Feibert *et al.* 2019; Silva-Aravena *et al.* 2020; Balkhi *et al.* 2022)

The anomaly-aware and ensemble-based demand forecasting architecture developed in this study aims to directly address these structural challenges. The results obtained are interpreted by jointly evaluating anomaly detection outputs, comparisons of parametric and non-parametric methods, time series structural analysis, and operational impacts.

Forecast Model Types Distribution

In this study, the demand forecasting process was evaluated not only based on forecast accuracy but also on which model types can be preferred under which data conditions. Health inventory consumption series exhibit a highly heterogeneous structure in terms of frequency of use, continuity, variability, and clinical dependency. This situation makes it difficult to apply a uniform forecasting approach for all materials and makes data adequacy a critical factor in model selection (Subramanian 2021; Balkhi *et al.* 2022).

The model type distribution results obtained in this context reveal that a significant portion of health inventory data lacks sufficient and continuous data for model training. Particularly for infrequently used, procedure-specific, or seasonally active medical consumables, the short length of the historical observation window limits the ability of forecasting models to learn meaningful patterns. The literature indicates that such irregular and intermittent demand series specific to the health sector pose a significant challenge for both statistical time series methods and machine learning-based approaches (Silva-Aravena *et al.* 2020; Subramanian 2021).

When examining cases where data adequacy is ensured, ensemble approaches are seen to be more prominent than individual models. Linear models (ridge, lasso) offer advantages in capturing low-variance and more regular components; while tree-based methods (gradient boosting, random forest, XGBoost) can more effectively model non-linear relationships, sudden demand spikes, and complex interactions. Bringing these different model families together contributes to balancing prediction errors and obtaining more generalizable results (Shaub 2020; Kim *et al.* 2023; Chien *et al.* 2023).

These findings show that, rather than searching for the "best single model" in healthcare inventory demand forecasting, a flexible, multi-model approach that can adapt to the data structure is more rational. The model type distribution reveals that the ensemble-based strategy adopted in the study is supported not only theoretically but also by the data.

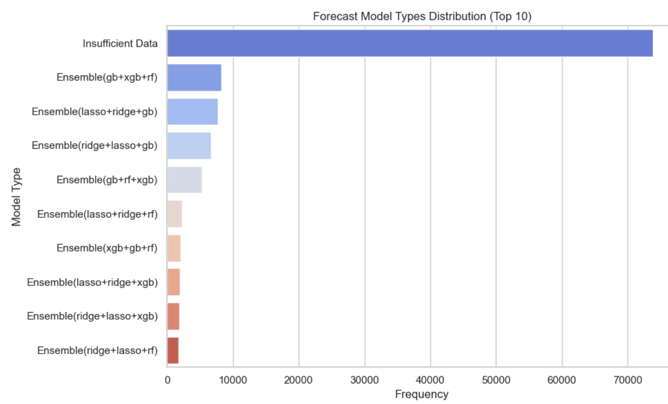

Figure 1 Distribution of the most commonly used prediction model types (Top 10).

As shown in Figure 1, one of the first outputs of the applied estimation process is the model distribution, which indicates which model types stand out in different material–unit pairs. Figure 1 presents the frequencies of the most commonly used estimation model types. The clear dominance of the "Insufficient Data" category in the graph indicates that a significant portion of the health inventory consumption data lacks sufficient and continuous observations for model training. In a hospital setting, some medical consumables are used infrequently, only in connection with specific clinical procedures, while others may be completely out of use during certain periods. This situation leads to short, irregular, and intermittent demand series, limiting the learning capacity of both statistical and machine learning-based models (Subramanian 2021; Balkhi *et al.* 2022).

When examining results outside the Insufficient Data category, it is observed that the most frequently preferred approaches are predominantly ensemble model combinations. This finding indicates that a single model family is insufficient to capture all behavioral patterns in health inventory demand series. Linear models (ridge, lasso) offer advantages in capturing regular and low-variance components, while tree-based methods (gradient boosting, random forest, XGBoost) can more effectively model non-linear relationships, sudden jumps, and interactions. Ensemble approaches, which combine these different strengths, produce more stable and generalizable predictions by balancing errors (Kim *et al.* 2023; Chien *et al.* 2023; Shaub 2020). Therefore, Figure 1 shows that the multi-model and flexible prediction strategy adopted in this study is also supported by the data.

Forecast Reliability Levels

Another factor as important as error metrics in evaluating forecast performance is the reliability level of the generated forecasts. Since forecast results directly influence operational decisions in healthcare inventory management, the question of how reliable the forecasts are is of critical importance. Particularly in demand series with high uncertainty, careful consideration should be given to how forecast outputs can be used for decision support (Goncalves *et al.* 2020). In this study, forecasts were classified under Very Low, Low, Medium, High, and Very High Reliability levels. The distribution obtained indicates that forecast uncertainty is high in a significant portion of healthcare inventory consumption data. Short data windows, high variance, sudden demand spikes, and irregular usage patterns widen the uncertainty intervals of forecast models, thereby reducing their reliability. The literature emphasizes that demand forecasting in the healthcare sector inherently involves high uncertainty and that this uncertainty must be explicitly managed (Subramanian 2021; Balkhi *et al.* 2022).

However, the presence of significant density in the Medium, High, and Very High Reliability categories indicates that some materials exhibit more regular and predictable consumption behavior. It has been observed that prediction reliability increases in series with more pronounced trend and seasonality components, lower coefficient of variation (CV), and higher observation continuity. In such series, ensemble approaches in particular are seen to produce more stable and reliable forecasts (Shaub 2020; Kim *et al.* 2023).This differentiation in forecast reliability levels reveals that a uniform forecasting strategy is not appropriate for all materials. While prediction outputs for materials in the Very Low Reliability group should be used only to a limited extent for decision support purposes, greater weight can be given to prediction-based automatic stock decisions for materials in the High and Very High Reliability groups. This approach is consistent with the literature arguing that prediction results in healthcare inventory management should be differentiated based on reliability (Goncalves *et al.* 2020).

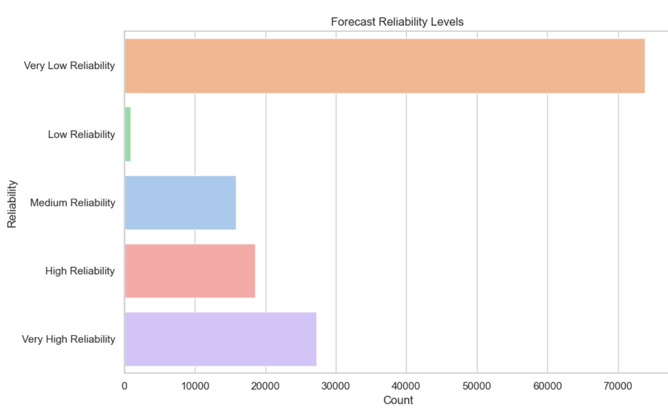

Figure 2 Distribution of prediction reliability levels.

Forecast performance has been evaluated not only through error metrics but also through the reliability levels of the forecasts. Figure 2 shows the distribution of the generated forecasts according to Very Low, Low, Medium, High, and Very High Reliability levels. The dominance of the Very Low Reliability category in the graph reveals that uncertainty is quite high in health inventory consumption series. Short data windows, high variance, sudden demand spikes, and irregular usage patterns widen the uncertainty intervals of prediction models, thereby reducing reliability (Subramanian 2021).Nevertheless, a significant concentration is also observed in the Medium, High, and Very High Reliability categories. This indicates that some materials have more regular consumption behaviour and that prediction models can produce more reliable outputs for these series. Prediction reliability increases in series where trend and seasonality components are more pronounced, the coefficient of variation (CV) is lower, and observation continuity is higher. The fact that ensemble approaches produce more successful results for materials in this group is consistent with the properties of reducing error variance and increasing generalisability emphasised in the literature (Shaub 2020; Kim *et al.* 2023).

One of the most important contributions of Figure 2 is that it shows that a single prediction strategy is not suitable for all materials. It is understood that prediction outputs should be used for decision support purposes only to a limited extent for materials in the Very Low Reliability group, whereas more weight should be given to prediction-based automatic stock decisions for materials in the High and Very High Reliability groups. This finding is consistent with studies arguing that prediction results should be differentiated based on reliability in healthcare inventory management (Goncalves *et al.* 2020; Balkhi *et al.* 2022).

Anomaly Detection Findings and Their Effects on the Health Inventory

The anomaly detection step applied prior to the forecasting process is a critical pre-processing stage in health inventory demand forecasting. The literature clearly states that feeding outliers and irregular observations directly into the model increases estimation errors, reduces model stability, and can lead to incorrect stock decisions (Ingle *et al.* 2021; Kim *et al.* 2023; Subramanian 2021; Merkuryeva *et al.* 2019).In this study, the DBSCAN algorithm is preferred for anomaly detection. Thanks to its density-based structure, DBSCAN can produce effective results in complex and heterogeneous data sets without the need for predefined threshold values or distribution assumptions.

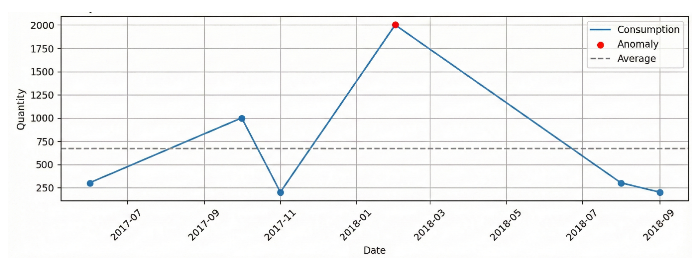

Figure 3 Singular and high-impact anomaly detected in the ADSH -1 Ground Floor Inventory and Biomedical Storage consumption series using the DBSCAN algorithm.

As observed in Figure 3, the DBSCAN algorithm clearly detected a high-volume singular anomaly in the consumption series belonging to the central storage facility. It can be seen that the consumption amount during the relevant period was significantly above the historical average. Such sudden and sharp consumption spikes are generally associated with urgent clinical needs, unexpected patient surges, or operational planning-out usage scenarios in the context of healthcare inventory (Feibert *et al.* 2019; Subramanian 2021).The literature emphasises that such singular but highly impactful anomalies create a disproportionate disruptive effect on prediction models. Particularly when error metrics have a squared structure (e.g., RMSE), such outliers can significantly degrade model performance (Merkuryeva *et al.* 2019; Ingle *et al.* 2021).Therefore, identifying and separating these anomalies prior to the prediction process facilitates the model's learning of the overall behaviour.

The most significant contribution of the DBSCAN algorithm in this example is its ability to successfully distinguish sudden changes in consumption intensity without relying on distribution assumptions or using fixed threshold values. This clearly demonstrates why density-based approaches are more suitable for heterogeneous and irregular data structures such as health inventories (Balkhi *et al.* 2022).

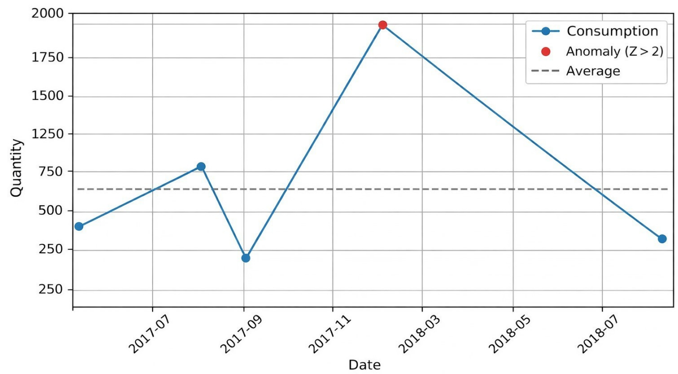

Figure 4 Multiple and irregular anomalies detected in the consumption series of the Güzeloba ADSP Laboratory unit using the DBSCAN algorithm

As shown in Figure 4, the consumption series of the laboratory unit contains multiple and irregular anomalies. Such consumption patterns are frequently observed in healthcare institutions because laboratory services are directly linked to patient profiles and clinical demand. Fluctuations in diagnostic and examination processes, in particular, can cause changes in consumption series intensity (Polater and Demirdogen 2018; Feibert *et al.* 2019).

The most important point to note in Figure 4 is that anomalies appear not only at a single point but at different time intervals and at different density levels. This explains why parametric methods based on fixed thresholds can be inadequate. The DBSCAN algorithm has successfully modelled this complex structure thanks to its ability to distinguish between different density regions. The literature indicates that consumption series for laboratory and imaging units in healthcare supply chains are typically multimodal and irregular (Balkhi *et al.* 2022; Subramanian 2021). In this context, Figure 4 visually supports why the DBSCAN-based approach should be preferred in complex systems such as healthcare inventory.

Comparative Evaluation of Parametric and Non-Parametric Anomaly Approaches

In order to evaluate the effectiveness of the DBSCAN-based approach more comprehensively, a comparative analysis was performed with the Z-Score, a classical parametric method. The Z-Score method is based on the assumption that the data follows an approximate normal distribution and identifies outliers using fixed threshold values.

Figure 5 shows the anomaly detection results obtained using the Z-Score method for the same central storage facility. Although the Z-Score method was able to detect significant outliers, it could only capture structural fluctuations and intensity changes in the consumption series to a limited extent. This situation stems from

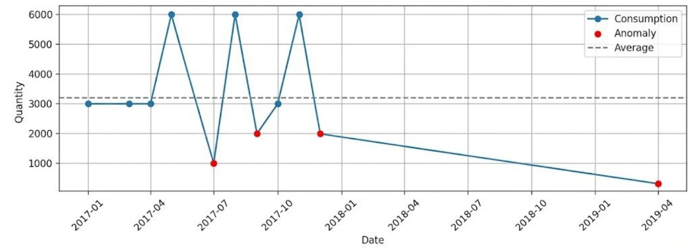

Figure 5 Anomalies detected in the ADSH -1 Ground Floor Inventory and Biomedical Storage consumption series using the Z-Score method.

the Z-Score's dependence on the distribution assumption. Health inventory consumption data often do not satisfy the normal distribution assumption; instead, they exhibit skewed, multimodal, and seasonally varying structures (Merkuryeva *et al.* 2019; Subramanian 2021).It is frequently emphasised in the literature that methods based on fixed thresholds in such data structures can only detect extreme observations but may overlook more complex anomalies (Ingle *et al.* 2021; Kim *et al.* 2023).

In this context, Figure 5 clearly illustrates the limitations of the Z-Score method in the health inventory context and supports why non-parametric approaches such as DBSCAN offer more flexible solutions (Silva-Aravena *et al.* 2020).

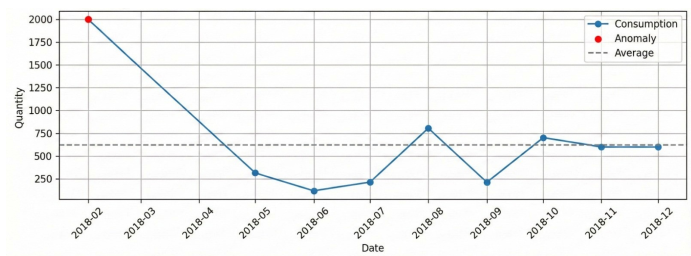

Figure 6 Results of anomaly detection performed on the consumption series of the ADSM A17 X-ray unit using the Z-Score method.

Figure 6 clearly shows that the Z-Score method can completely miss anomalies in some periods. The consumption series for the X-ray unit exhibits structural breaks over time due to changing clinical demands and patient profiles. The performance of methods based on fixed thresholds is significantly reduced in such series.

The literature indicates that consumption series for imaging and diagnostic units typically contain seasonality, trends, and sudden increases in demand (Feibert *et al.* 2019; Balkhi *et al.* 2022). In this context, Figure 4 illustrates why parametric methods are limited in the healthcare inventory context and why approaches based on distribution assumptions do not always produce reliable results. This finding directly aligns with studies explaining why non-parametric methods and density-based approaches are increasingly preferred in the healthcare supply chain literature (Merkuryeva *et al.* 2019; Subramanian 2021).

Following anomaly cleaning, the time series decomposition method was applied to enable a more in-depth analysis of the fundamental behavioural characteristics of the consumption series. This analysis allows for the separate examination of the trend, seasonality, and residual components.

Figure 7 shows the time series decomposition results obtained for material 17380. While a long-term increasing structure is ob-

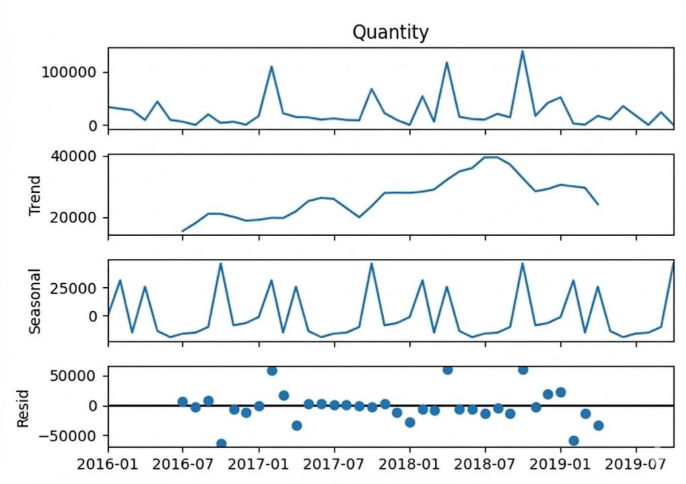

Figure 7 Time series decomposition analysis for material 17380: trend, seasonality and residual components.

served in the trend component, pronounced fluctuations are noticeable in the seasonal component. This indicates that the consumption of the relevant material exhibits both an increasing trend over time and is influenced by seasonal factors. The irregularities observed in the residual components reveal that this series has a structure susceptible to anomalies. The literature emphasises that single forecasting models are generally inadequate for such complex time series, and that ensemble approaches produce more successful results (Kim *et al.* 2023; Chien *et al.* 2023; Sina *et al.* 2023; Shaub 2020).

In this context, Figure 5 is one of the key figures that visually supports why ensemble and hybrid forecasting approaches are necessary.

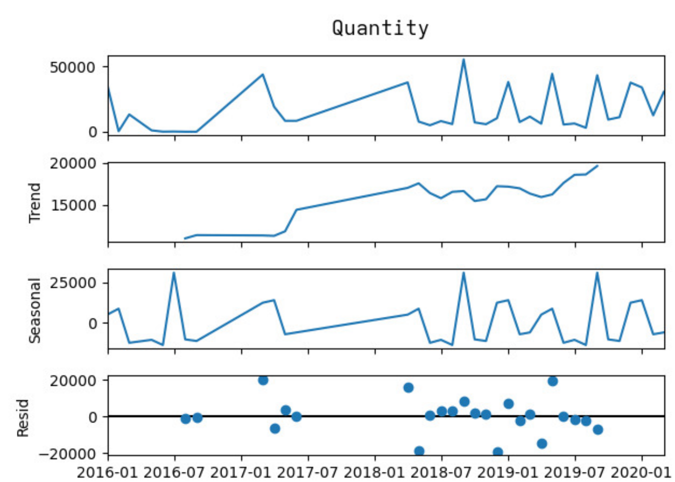

Figure 8 Time series decomposition analysis for material 17305: trend, seasonality and residual components.

Figure 8 presents the decomposition results obtained for a material with a high-volume and complex consumption structure. In addition to the trend and seasonality components, the high variance observed in the residual components indicates that this series is both prone to anomalies and difficult to forecast. The literature indicates that such high-volume and irregular series cannot be effectively predicted using simple statistical models; instead, multiple model combinations provide more stable results (Sharma *et al.* 2021; Sina *et al.* 2023). Therefore, Figure 6 is of critical importance in justifying the proposed ensemble-based approach.

CONCLUSION

The findings obtained within the scope of this study not only present results related to statistical prediction accuracy but also demonstrate that the developed anomaly-aware and ensemble-based prediction architecture produces outputs that can be directly integrated into the operational and strategic decision-making processes of healthcare institutions. As frequently emphasised in the literature, healthcare inventory data has an irregular, sparse, highly variable, and anomaly-prone structure necessitated the design of the developed system to directly address these structural challenges (Merkuryeva *et al.* 2019; Subramanian 2021; Feibert *et al.* 2019; Silva-Aravena *et al.* 2020; Balkhi *et al.* 2022).The anomaly-aware forecasting process enables unexpected demand spikes to be detected at an early stage, contributing to reducing the risk of stock depletion, preventing excessive inventory costs, and ensuring service continuity (Karamshetty *et al.* 2022; Goncalves *et al.* 2020). The generated forecasts are not only reported as point values but are also presented with an explainable reliability score calculated based on indicators such as the amount of data, behavioural stability, timeliness, and seasonal complexity for each material. Thus, decision-makers can obtain a quantitative answer not only to the question ‘how much will we consume?’ but also to the question ‘to what extent can I trust this forecast?’ (Subramanian 2021).

Situation indicates that a uniform stock and forecasting policy for all materials will not be effective and necessitates the integrated use of ABC classification, commonly used in classical inventory literature, and XYZ classification, which is based on demand predictability (Aktunc *et al.* 2019; Babai *et al.* 2015; Demiray Kirmizi *et al.* 2024).Thanks to the combined ABC–XYZ segmentation developed within the scope of the study, while stock removal or alternative usage evaluations are recommended for materials exhibiting low and irregular consumption, more protective policies such as increasing safety stock and redefining reorder points are systematically developed for critical materials exhibiting high variation. In cases where the data length is insufficient or the consumption pattern cannot be modelled in a statistically reliable manner, a conservative approach prioritising expert assessment is adopted instead of automatic decision-making. This holistic structure enables the evolution of forecasting outputs into actionable, traceable, and justifiable decision recommendations, demonstrating the practical application of integrated forecasting–classification approaches advocated in the literature for healthcare inventory management (Feibert *et al.* 2019; Subramanian 2021).

Ethical standard

The authors have no relevant financial or non-financial interests to disclose.

Availability of data and material

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

LITERATURE CITED

Abdul-Rahman, S., N. H. Zulkifley, I. Ibrahim, and S. Mutalib, 2021 Advanced machine learning algorithms for house price prediction: Case study in Kuala Lumpur. Science and Information Organization .

Adedunjoye, A. S. and J. O. Enyejo, 2024 Leveraging predictive analytics to improve demand forecasting and inventory management in healthcare supply chains. International Journal of Scientific Research in Science Engineering and Technology .

Aifuwa, S. E., T. Oshoba, E. Ogbuefi, P. N. Ike, S. B. Nnabueze, *et al.*, 2020 Predictive analytics models enhancing supply chain demand forecasting accuracy and reducing inventory management inefficiencies. International Journal of Multidisciplinary Research and Growth Evaluation .

Aktunc, E. A., A. Atamturk, and E. Demir, 2019 Demand classification and inventory control in healthcare systems using ABC–XYZ analysis. European Journal of Operational Research **274**: 539–551.

Babai, M. Z. *et al.*, 2015 Demand forecasting and inventory control in healthcare: A review. International Journal of Production Economics **167**: 130–146.

Balkhi, Z. *et al.*, 2022 Healthcare supply chain management: A review of quantitative models. Operations Research for Health Care **33**: 100349.

Cao, Y. and L. Gui, 2019 Multi-step wind power forecasting model based on LightGBM. In *Proceedings of the International Conference on Signal and Artificial Intelligence (ICSAI 2018)*, pp. 192–197.

Chien, C.-F., Y.-J. Chen, and J.-T. Peng, 2023 Hybrid and ensemble forecasting models for complex demand patterns. International Journal of Forecasting **39**: 287–302.

Dachepalli, V., 2025 An ARAS-based evaluation of AI applications for demand forecasting and inventory management in supply chains. International Journal of Cloud Computing and Supply Chain Management .

Darshan, U., G. P. Chinmaya, H. Abhinav, S. V. Kumar, R. Deepak, *et al.*, 2025 Predictive analysis on medicines availability in hospitals using machine learning and deep learning technique. International Journal For Multidisciplinary Research .

Demiray Kirmizi, M. *et al.*, 2024 Integrated inventory classification and forecasting in healthcare supply chains. Journal of Manufacturing Systems **72**: 102–118.

Donkor, A. A., S. A. Dada, A. Korang, and J. Umoren, 2024 The role of artificial intelligence and machine learning in optimizing U.S. healthcare supply chain management. World Journal of Advanced Research and Reviews .

Erdebilli, B. and B. Devrim-Tenba, 2022 Ensemble voting regression based on machine learning for predicting medical waste: A case from Turkey. Multidisciplinary Digital Publishing Institute .

Fatima, S. S. W. and A. Rahimi, 2024 A review of time-series forecasting algorithms for industrial manufacturing systems. Multidisciplinary Digital Publishing Institute .

Feibert, D. C., R. K. Hansen, and P. Jacobsen, 2019 Measuring supply chain resilience in healthcare: An empirical study. Supply Chain Management: An International Journal **24**: 78–96.

Gldoan, E., F. H. Yan, A. Pinar, C. Olak, S. Kadry, *et al.*, 2023 A proposed tree-based explainable artificial intelligence approach for the prediction of angina pectoris. Nature Portfolio .

Goncalves, J. N. *et al.*, 2020 Decision support systems for healthcare inventory management. Computers & Operations Research **113**: 104785.

Gurumurthy, A., V. K. Nair, and S. Vinodh, 2021 Application of a hybrid selective inventory control technique in a hospital. The TQM Journal **33**: 568–595.

Ingle, V., P. Shah, and A. Shukla, 2021 Time series forecasting for healthcare demand using hybrid models. Expert Systems with Applications **184**: 115555.

Jahin, M. A., A. Shahriar, and M. A. Amin, 2024 MCDFN: Supply chain demand forecasting via an explainable multi-channel data fusion network model. Evolutionary Intelligence .

Joshi, S. V., S. B. Patil, S. N. Patil, S. Bhosle, N. Karhadkar, *et al.*, 2025 Machine Learning Applications in Hospital Pharmacy for Predicting Drug Shortages and Supply Chain Optimization. Research Journal of Pharmacy and Technology .

Karamshetty, S., R. Akkiraju, and K. Bhaduri, 2022 Anomaly detection in supply chain demand using density-based approaches. IEEE Access **10**: 41245–41257.

Khalid, O., 2024 Short-term and long-term product demand forecasting with time series models. Journal of Trends in Financial and Economics .

Kim, S., J. Park, and D. Lee, 2023 Ensemble learning-based demand forecasting for healthcare inventory management. Applied Soft Computing **132**: 109840.

Kolambe, S. A. M., 2024 Forecasting the future: A comprehensive review of time series prediction techniques. None .

Kontopoulou, V. I., A. D. Panagopoulos, I. Kakkos, and G. K. Matsopoulos, 2023 A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks. Multidisciplinary Digital Publishing Institute .

Luo, L., L. Luo, X. Zhang, and X. He, 2017 Hospital daily outpatient visits forecasting using a combinatorial model based on ARIMA and SES models. BMC Health Services Research .

Mbonyinshuti, F., J. Nkurunziza, J. Niyobuhungiro, and E. Kayitare, 2022 The prediction of essential medicines demand. Processes **10**: 26.

Mbonyinshuti, F., J. Nkurunziza, J. Niyobuhungiro, and E. Kayitare, 2024 Health supply chain forecasting: A comparison of ARIMA and LSTM time series models for demand prediction of medicines. Acta Logistica .

Merkuryeva, G., A. Valberga, and A. Smirnov, 2019 Demand forecasting in healthcare supply chains: A systematic literature review. Procedia Computer Science **149**: 541–550.

Motamedi, M., J. W. Dawson, N. Li, D. G. Down, and N. M. Heddle, 2024 Demand forecasting for platelet usage: From univariate time series to multivariable models. Public Library of Science .

Pawlikowski, M. and A. Chorowska, 2020 Weighted ensemble of statistical models. International Journal of Forecasting **36**: 93–97.

Perone, G., 2021 Comparison of ARIMA, ETS, NNAR, TBATS and hybrid models to forecast the second wave of COVID-19 hospitalizations in Italy. Springer Science+Business Media .

Polater, A. and G. Demirdogen, 2018 Healthcare logistics and demand uncertainty: A case-based analysis. Journal of Health Systems **4**: 45–59.

Punnahitanond, W., P. Kantavat, N. Prompoon, and T. Ananpiriyakul, 2025 Forecasting drug dispensation using machine learning: A case study on a large public hospital. In *2025 8th International Conference on Artificial Intelligence and Big Data (ICAIBD)*.

Qiu, H., L. Luo, Z. Su, L. Zhou, L. Wang, *et al.*, 2019 Machine learning approaches to predict peak demand days of cardiovascular admissions considering environmental exposure. BMC Medical Informatics and Decision Making .

Sharma, A., R. Gupta, and J. Saini, 2021 Machine learning approaches for demand forecasting: A review. Materials Today: Proceedings **46**: 11131–11138.

Shaub, M., 2020 Why ensemble learning improves forecasting accuracy. Journal of Forecasting **39**: 38–55.

Shern, S. J., M. T. Sarker, M. H. S. M. Haram, G. Ramasamy, S. P. Thiagarajah, *et al.*, 2024 Artificial intelligence optimization for user prediction and efficient energy distribution in electric vehicle smart charging systems. Energies .

Shih, H. and S. Rajendran, 2019 Comparison of time series methods and machine learning algorithms for forecasting Taiwan Blood Services Foundation's blood supply. Hindawi Publishing Corporation .

Silva-Aravena, M., R. Morales, and L. Cisternas, 2020 Forecasting healthcare demand using time series and machine learning models. Computers & Industrial Engineering **143**: 106419.

Sina, M., M. Rahimi, and R. Tavakkoli-Moghaddam, 2023 A hybrid machine learning approach for demand forecasting in healthcare supply chains. Computers & Industrial Engineering **176**: 108933.

Subramanian, N., 2021 Healthcare inventory management: A review of demand forecasting and decision support systems. International Journal of Production Economics **231**: 107867.

Umoren, J., T. O. Agbadamasi, T. K. Adukpo, and N. Mensah, 2025 Leveraging artificial intelligence in healthcare supply chains: Strengthening resilience and minimizing waste. EPRA International Journal of Economics, Business and Management Studies .

Vignesh, A. and N. Vijayalakshmi, 2025 A weighted ensemble model combining ARIMA, LSTM, and GBM for robust time series prediction. Scientific and Technical Journal of Information Technologies, Mechanics and Optics .

How to cite this article: Özkurt, C., Küçükler, A. K., Karslıoğlu, M., and Özdemir, R. N. Enhancing Hospital Inventory Forecasting Accuracy through Hybrid and Ensemble Learning Models. *Computers and Electronics in Medicine*, 3(1), 60-76, 2026.