Comparison of Time Series Methods and Machine Learning Algorithms for Forecasting Taiwan Blood Services Foundation’s Blood Supply

Han Shih¹ and Suchithra Rajendran²

¹Department of Industrial and Manufacturing Systems Engineering, University of Missouri, Columbia, MO 65211, USA
²Department of Marketing, University of Missouri, Columbia, MO 65211, USA

Correspondence should be addressed to Suchithra Rajendran; rajendrans@missouri.edu

Received 17 July 2019; Accepted 27 August 2019; Published 17 September 2019

Academic Editor: Feng-Huei Lin

Copyright © 2019 Han Shih and Suchithra Rajendran. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Purpose. The uncertainty in supply and the short shelf life of blood products have led to a substantial outdating of the collected donor blood. On the other hand, hospitals and blood centers experience severe blood shortage due to the very limited donor population. Therefore, the necessity to forecast the blood supply to minimize outdating as well as shortage is obvious. This study aims to efficiently forecast the supply of blood components at blood centers. Methods. Two different types of forecasting techniques, time series and machine learning algorithms, are developed and the best performing method for the given case study is determined. Under the time series, we consider the Autoregressive (AUTOREG), Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA), Seasonal ARIMA, Seasonal Exponential Smoothing Method (ESM), and Holt-Winters models. Artificial neural network (ANN) and multiple regression are considered under the machine learning algorithms. Results. We leverage five years worth of historical blood supply data from the Taiwan Blood Services Foundation (TBSF) to conduct our study. On comparing the different techniques, we found that time series forecasting methods yield better results than machine learning algorithms. More specifically, the least value of the error measures is observed in seasonal ESM and ARIMA models. Conclusions. The models developed can act as a decision support system to administrators and pathologists at blood banks, blood donation centers, and hospitals to determine their inventory policy based on the estimated future blood supply. The forecasting models developed in this study can help healthcare managers to manage blood inventory control more efficiently, thus reducing blood shortage and blood wastage.

1. Introduction

Blood performs several important functions in the human body such as transporting oxygen, carrying supplements to our cells, disposing ammonia, carbon dioxide, and other waste items. Four of the most critical elements are the red blood cells (RBC), white blood cells (WBC), plasma, and platelets [1]. The American Red Cross reported that over 35,000 RBC units, 10,000 plasma units, and 7,000 platelet units are required day-to-day within the US [2]. Due to the short shelf life of blood components, hospitals and blood centers are faced with the challenge of maintaining appropriate inventory levels to avoid outdating and shortage.

Managing blood supply and demand is the core part of the healthcare supply chain system as blood plays a very crucial role in saving human lives. Blood supply forecasting is essential for making supply chain decisions, such as donor drive scheduling, vehicle routing policies, and inventory management, at blood centers and hospitals. Accurate forecasts of the timing and amount of future blood requests have been considered as the key inputs to donor recruitment decision making and inventory control. It is important to gather data for several years to forecast monthly demand and to recognize seasonality in demand [3–6]. Lestari et al. [7] indicated that the forecasting can predict the data trend observed and future demand for blood components.
2. Literature Review

Several studies have leveraged time series forecasting techniques for predicting the blood demand at hospitals and blood centers. For instance, Pereira [8] investigated and evaluated the autoregressive integrated moving average (ARIMA) model and Holt-Winters exponential smoothing model to predict monthly demand for red blood cell transfusions at a tertiary care. While these methods focused on using time series forecast, Bosnes et al. [9] used the statistical regression technique for the forecast of blood donor arrivals at the blood bank of Oslo and found that the most important factors among 18 explanatory variables were: donor age, time from making an appointment to arriving at the drive, contact methods used, number of prior donations, and donor no-show rate. Fortsch and Khapalova [10] introduced numerous practical methods to predict future demand of blood. Several forecasting models, including the naïve, exponential smoothing, moving average, and time series decomposition, were tested using the daily demand data from a blood center that were obtained for January 2006 to December 2012. They also compared the performance of these methods with an autoregressive moving average (ARMA) model. The results revealed that the ARMA forecasting model performed better for eight out of nine time series model settings. Similarly, Khaldi et al. [11] explored the capabilities of employing machine learning algorithms such as the artificial neural network (ANN) model to predict future demand for blood.

3. Materials and Methods

As discussed earlier, the study aims to develop effective forecasting methods to predict the supply of RBCs using two different techniques: time series forecasting methods and machine learning algorithms.

3.1. Time Series Forecasting

This section discusses the seven time series forecasting methods used in this study.

3.1.1. Autoregressive (AUTOREG) Model [12, 13]. The AUTOREG procedure estimates and forecasts linear regression models for time series data when the errors are autocorrelated. The autoregressive model regresses the value of the series at time \( t \) \((Y_t)\) on the values during the time periods \( t-1, t-2, \ldots, t-p \). The mathematical formula is expressed as follows:

\[
Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + \alpha_p Y_{t-p} + \epsilon_t,
\]

where \( \alpha_0, \alpha_1, \ldots, \alpha_p \) are the linear regression coefficients, \( Y_t \) is the forecasted value at time \( t \), and \( \epsilon_t \) is the random error variable and is generally assumed to have a normal distribution with mean 0 and variance \( \sigma^2 \) (i.e., normal \((0, \sigma^2)\)).

3.1.2. Autoregressive Moving Average (ARMA) Models [12–14]. ARMA model is one of the basic tools in time series modeling. Suppose the time series \( Y_1, Y_2, \ldots, Y_t \) is a stationary stochastic process time series, the expression ARMA

\( (p, q) \) represents the model with autoregressive order of \( p \) and moving-average order of \( q \). This model is a combination of the AR \( (p) \) and MA \( (q) \) models, where AR \( (p) \) is written as

\[
Y_t = a + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \epsilon_t,
\]

and MA \( (q) \) is written as

\[
Y_t = b + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t.
\]

3.1.3. Autoregressive Integrated Moving Average (ARIMA) Model [12–14]. The ARIMA (autoregressive integrated moving average) approach was made popular by Box-Jenkins models [11]. The ARIMA procedure is functioning as a linear combination of its current values, past values, past errors, and past values of other time series (predictor time series) to predict a future response value in a time series.

With time series nonstationary behavior, the above ARMA \((p, q)\) model can be extended and written using difference which is defined as:

\[
Y_t - Y_{t-1} = (1 - B)Y_t = \nabla Y_t,
\]

where \( t \) is the index of time, \( Y_t \) is time series \( \{Y_t: 1 \leq t \leq n\} \) at time \( t \), and \( B \) is the backward shift operator, which means that \( B \) has the effect of shifting the data back one period (i.e., \( B Y_t = Y_{t-1} \)).

3.1.4. Seasonal ARIMA Model [12, 13, 15, 16]. Seasonal ARIMA model is written with the general expression ARIMA \((p, d, q)(P, D, Q)\). The symbol \( p \) is the order of the nonseasonal autoregressive component, \( d \) is the order of the differencing, \( q \) is the order of the nonseasonal moving-average process, \( P \) is the order of the seasonal autoregressive part, \( D \) is the order of the seasonal differencing, \( Q \) is the order of the seasonal moving-average process, and \( s \) is the duration of the seasonal cycle.

Let \( Y_t \) be a dependent time series \( \{Y_t: 1 \leq t \leq n\} \) at time \( t \), then the mathematical formula for the seasonal ARIMA model is expressed as follows:

\[
(1 - B)^d (1 - B^s)^D Y_t = \mu + \theta(B)\phi(B)\epsilon_t,
\]

where \( \mu \) is the constant mean, \( B^s \) is the seasonal backward shift operator, \( \phi(B) = 1 - \phi_1 B - \cdots - \phi_s B^s \) is the seasonal autoregressive component, and \( \theta(B) = 1 - \theta_1 B^s - \cdots - \theta_s B^{sQ} \) is the seasonal moving-average component.

3.1.5. Seasonal Exponential Smoothing Model [12, 13, 15, 16]. In the seasonal exponential smoothing method (ESM), the equation of forecast value at time \( t + k(Y_{t+k}) \) is given by
The smoothing equations are as follows:

\[ L_t = \alpha (X_t - S_{t-p}) + (1 - \alpha) L_{t-1}, \]

\[ S_t = \gamma (X_t - L_t) + (1 - \gamma) S_{t-p}, \]

where \( X_t \) is given observation at time \( t \), and \( \alpha \) and \( \gamma \) are the level and seasonal smoothing parameters, respectively, \( L_t \) is the estimated level component at time \( t \), \( S_t \) is the estimated seasonal component at time \( t \), and \( p \) is the periods after which the seasonal cycle repeats itself.

### 3.1.6. Multiplicative Holt-Winters Model [12, 13, 15, 16].

The Holt-Winters model, also known as the triple exponential smoothing, applies three types of exponential smoothing to the time series—value, trend, and seasonality. The model equation for the Holt-Winters method can be either additive or multiplicative model. In this section, we present the multiplicative Holt-Winters Model. In this section, we present the multiplicative Holt-Winters Model, whereas Section 3.1.7 presents the additive model.

The mathematical formula relevant to a time series with a trend and constant seasonal component using the Holt-Winters additive technique has the forecast at time \( t + k \) \( Y_{t+k} \) given by following equation:

\[ Y_{t+k} = (L_t + kT_t)S_{t+k-p}. \]

The smoothing equations are given using the following equations:

\[ L_t = \alpha \left( \frac{X_t}{S_{t-p}} \right) + (1 - \alpha) (L_{t-1} + T_{t-1}), \]

\[ T_t = \beta (L_t - L_{t-1}) + (1 - \beta) T_{t-1}, \]

\[ S_t = \gamma \left( \frac{X_t}{T_t} \right) + (1 - \gamma) S_{t-p}, \]

where \( X_t \) is given observation at time \( t \), \( \alpha \), \( \beta \), and \( \gamma \) are the level, trend, and seasonal smoothing parameters, respectively, \( L_t \) is the estimated level component at time \( t \), \( T_t \) is the estimated trend component at time \( t \), \( S_t \) is the seasonality index at time \( t \), and \( p \) is the periods after which the seasonal cycle repeats itself.

### 3.1.7. Additive Holt-Winters Model [12, 13, 15, 16].

In this section, we present the additive Holt-Winters Model.

For the additive model, the forecasted supply estimate for time \( t + k \) is given by the following equation:

\[ Y_{t+k} = L_t + kT_t + S_{t-p+k}. \]

The estimates of level, trend, and seasonal factors for additive model equations are given using the following equations:

\[ L_t = \alpha (Y_t - S_{t-p}) + (1 - \alpha) (L_{t-1} + T_{t-1}), \]

\[ T_t = \beta (L_t - L_{t-1}) + (1 - \beta) T_{t-1}, \]

\[ S_t = \gamma (Y_t - L_t) + (1 - \gamma) S_{t-p}. \]

3.2. Machine Learning Algorithms. Machine learning is a technology exploring the algorithms to analyze a set of data, learn from the insights gathered, and make predictions on data [17]. For the blood supply forecasting, we leverage the two most widely used machine-learning techniques, artificial neural network and regression.

3.2.1. Artificial Neural Networks (ANN). ANN is a reinforcement learning method that is an adaptation of biological neural network. The network consists of several nodes that are distributed across numerous layers, and each layer is connected to its previous and subsequent layers within the network [17]. These interconnected elements work closely to process information that they receive from the nodes of the previous layers and transfer them to the next layer based on the sigmoid function. They are particularly useful for modeling complex relationships in high-dimensional data or where the relationship between the input and output variables is not easy to understand [17].

3.2.2. Multiple Regression. Multiple regression is another class of problem in machine learning that is trying to predict a continuous value of a variable instead of a class unlike classification problem [17]. Linear regression with ordinary least square is one of the classic machine learning algorithms in this domain. The mathematical formula for the regression model is represented as follows:

\[ Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n + \epsilon, \]

where \( Y \) is the response variable, \( X_i \) is an independent variable, \( \beta_0 \) is the intercept, \( \beta_i \) is the slope of the coefficient \( X_i \) (both \( \beta_0 \) and \( \beta_i \) are unknown coefficients to be estimated by the model), and \( \epsilon \) is the error variable.

3.3. Evaluation of the Different Methods. We use four different measures of forecast errors for evaluating the model performance and the accuracy of the methods; they are MAE, MSE, BIAS, and MAPE [12, 15, 18].

Assume \( X_1, X_2, \ldots, X_n \) are actual data and \( F_1, F_2, \ldots, F_n \) are forecasted data, and then the \( n \) values of forecast errors, \( e_1, e_2, \ldots, e_n \), are given by \( e_i = F_i - X_i \), \( e_2 = F_2 - X_2, \ldots, e_n = F_n - X_n \).

(a) Mean absolute error (MAE): it measures the average significance of the forecast errors, where all individual errors have equal weights:

\[ MAE = \frac{1}{n} \sum_{i=1}^{n} |e_i|. \]

(b) Mean squared error (MSE): it also measures the significance of the forecast errors, and larger errors get penalized more due to squaring:
\[ \text{MSE} = \frac{1}{n} \sum_{i=1}^{n} e_i^2. \]  
\[ \text{BIAS} = \sum_{i=1}^{n} e_i. \]  

(c) BIAS: this is an indication of whether the forecast is overestimating or underestimating the actual supply over the forecast horizon: 

(d) Mean absolute percentage error (MAPE): it measures the relative significance of forecasting errors in percentage terms:
$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{e_i}{X_i} \right| \times 100. \quad (19)$$

4. Results

4.1. Data Collection. The historical supply data for five years from 2013 to 2017 are first gathered from the health records. The summary statistics are given in Table 1.

From Table 1, it is observed that the average blood supplies of the weekdays for each year are steady. Also, we can see that Monday supply is very high, Thursday and Friday supplies are quite high, Tuesday and Wednesday supplies are moderate, and Saturday and Sunday supplies are significantly lower.

4.2. Time Series Forecasting Results. After running the seven different time series models discussed in Section 3.1 and obtaining the forecasts, we evaluate them using the error measures given in Section 3.3, and the results are presented in Table 2. It is clear that Seasonal ARIMA Model, Seasonal Exponential Smoothing Method, and Multiplicative Holt-Winters Model yield minimal error measures. Hence, we conclude that, under the time series methods, these three models are best forecasting the blood supply for the case study data under consideration.

4.3. Machine Learning Algorithm Results. The performance of the machine learning algorithms is compared in Table 3. For this particular dataset, results show that regression is a better predictor of the blood supply, nevertheless, the power of the results using regression is quite low ($R^2 = 63.71\%$).

Therefore, regression is used to predict the supply for the first week of January 2018 as shown in Table 4. A summary of the results obtained under the time series method and regression is given in Table 4.

Clearly from the results, we can infer that there is not a single method that predicts the supply accurately, and hence we recommend using the average value of the forecasts obtained under these four methods for estimating the future supply [15, 19–21].

5. Discussion

This study focusses on predicting the supply of red blood cells for Taiwan Blood Services Foundation (TBSF) [22], a nongovernmental and nonprofit organization. So far, more than seven million citizens have donated blood in Taiwan through this foundation (which accounts for over 25% of the total population of Taiwan) [23]. Currently, blood centers at TBSF do not have a proper blood forecasting system, and some blood centers face blood shortage problems as a result to lack of accurate forecasting of blood supply. This paper focusses on developing a blood supply forecasting decision support tool for TBSF using time series and machine learning algorithms. The accurate forecasting models will enable TBSF to make good blood supply chain management planning decisions, such as when to collect blood from donors, how much units to collect, proper assignment of the workforce for collecting blood in donor drives, and blood component testing process. Upon accurately forecasting the future supply using the methods discussed in this study, inventory models can then be developed to make decisions on the number of units to order and time between orders.

There are some limitations on forecasting methods. Accuracy of forecasting could be affected by various factors. If there are some unknown variable(s) that could cause some of the fluctuations in the data, then it will be more difficult to forecast unless there are known explanatory variable(s) accounting for the variations. Blood supply forecasting is vital for blood supply chain decisions, and they have to be updated as more reliable information becomes available. Hence, after appropriate forecasting methods are selected, it is important to continuously monitor the forecast accuracy.

Data Availability

The data used to support the findings of this study have not been made available because they are confidential to the case study blood center and hospitals.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.
Acknowledgments

We are grateful to Kuan-Tsou (Johnny) Lin, Director of Department of Operation, and Ming Chang Lin, Director of Hsin Chu Blood Center at the TBSF, for providing us with five years of daily blood supply data. We would also like to show our gratitude to Sabrina Lei Li, Director of Department of Public Relations, who provides important insight and expertise that greatly assisted the research. The first author is grateful to the US Department of Education for funding his PhD study through the Graduate Assistance in Areas of National Need (GAANN) fellowship.

References

[1] American Society of Hematology, Blood Basics, American Society of Hematology, Washington, DC, USA, 2018, http://www.hematology.org/Patients/Basics.
[2] American Red Cross, Blood Types, American Red Cross, Washington, DC, USA, 2018, https://www.redcrossblood.org/learn-about-blood/blood-types.html.
[3] W. P. Pierskalla, "Supply chain management of blood banks," in Operations Research and Health Care: A Handbook of Methods and Applications, Kluwer Academic Publishers, pp. 103–145, New York, NY, USA, 2014.
[4] S. Rajendran and A. R. Ravindran, "Platelet ordering policies at hospitals using stochastic integer programming model and heuristic approaches to reduce wastage," Computers & Industrial Engineering, vol. 110, pp. 151–164, 2017.
[5] S. Rajendran and A. R. Ravindran, "Inventory management of platelets along blood supply chain to minimize wastage and shortage," Computers & Industrial Engineering, vol. 130, pp. 714–730, 2019.
[6] S. Srinivas and A. R. Ravindran, "Systematic review of opportunities to improve outpatient appointment systems," in Proceedings of the IIE Annual Conference, pp. 1697–1702, Institute of Industrial and Systems Engineers (IISE), Pittsburgh, PA, USA, 2017.
[7] F. Lestari, U. Anwar, N. Nugraha, and B. Azwar, "Forecasting demand in blood supply chain (case study on blood transfusion unit)," in Proceedings of the World Congress on Engineering, vol. vol II, London, UK, July 2017.
[8] A. Pereira, "Performance of time-series methods in forecasting the demand for red blood cells transfusion," Transfusion, vol. 44, no. 5, pp. 739–746, 2004.
[9] V. Bosnes, M. Aldrin, and H. E. Heier, "Predicting blood donor arrival," Transfusion, vol. 45, no. 2, pp. 162–170, 2005.
[10] S. M. Fortsch and E. A. Khapalova, "Reducing uncertainty in demand for blood," Operations Research for Health Care, vol. 9, pp. 16–28, 2016.
[11] R. Khalidi, A. E. Afia, R. Chihieb, and R. Faizi, Artificial Neural Network Based Approach for Blood Demand Forecasting: Fez Transfusion Blood Center Case, Mohammed V University, Rabat, Morocco, 2017.
[12] S. Nahmias, Production and Operations Analysis, McGraw-Hill, Irwin, CA, USA, 6th edition, 2008.
[13] SAS, Forecasting Process Details” Retrieved from SAS/ETS 14.3 User’s Guide, SAS Institute Inc., pp. 4150–4177, Cary, NC, USA, 2017.
[14] A. Pankratz, Forecasting with Univariate Box-Jenkins Models: Concepts and Cases, John Wiley & Sons, New York, NY, USA, 1983.
[15] A. Ravindran and D. P. Warsing, Supply Chain Engineering: Models and Applications, CRC Press, Boca Raton, FL, USA, 2013.
[16] R. J. Hyndman and G. Athanasopoulos, Forecasting Principles and Practice, O’Texts, Australia, 2nd edition, May 2018.
[17] S. Srinivas and S. Rajendran, "A data-driven approach for multiobjective loan portfolio optimization using machine-learning algorithms and mathematical programming," in Big Data Analytics Using Multiple Criteria Decision-Making Models, CRC Press, pp. 175–210, Boca Raton, FL, USA, 2017.
[18] S. Chopra and P. Meindl, Supply Chain Management: Strategy, Planning, and Operation, Pearson-Prentice Hall, Upper Saddle River, NJ, USA, 6th edition, 2015.
[19] P. H. Frances, “Averaging model forecasts and expert forecasts: why does it work,” Interface, vol. 41, no. 2, pp. 177–181, 2011.
[20] M. Gahirwal and M. Vijayalakshmi, Inter Time Series Sales Forecasting, Society’s Institute of Technology, Chembur, India, 2013, https://arxiv.org/abs/1303.0117.
[21] S. Rajendran, “Finite and infinite time horizon inventory models to minimize platelet wastage at hospitals,” International Journal of Operations and Quantitative Management, vol. 22, no. 2, pp. 119–140, 2016.
[22] Taiwan Blood Services Foundation, “Annual report,” 2018, http://intra.blood.org.tw/upload/cf15c0af-f84d-4628-9f8f-9661b6c34b8.pdf.
[23] Blood Donation Services in Taiwan, 2019, http://intra.blood.org.tw/upload/ec805ca8-9f64-45e3-ab07-20673a31164c.pdf.