



Contents lists available at ScienceDirect

## International Journal of Forecasting

journal homepage: [www.elsevier.com/locate/ijforecast](http://www.elsevier.com/locate/ijforecast)

# Machine learning algorithms for forecasting and backcasting blood demand data with missing values and outliers: A study of Tema General Hospital of Ghana

Clement Twumasi<sup>a,\*</sup>, Juliet Twumasi<sup>b</sup><sup>a</sup> School of Mathematics, Cardiff University, Abacus, Cardiff, CF24 4AG, UK<sup>b</sup> Ghana Institute of Management and Public Administration (GIMPA) Business School, P. O. Box AH 50, Achimota, Ghana

## ARTICLE INFO

## Keywords:

Blood demand  
Blood supply  
Forecasting  
Backcasting  
Kalman smoothing  
Imputation  
Machine learning  
Neural networks  
Time-reversibility

## ABSTRACT

The major challenge in managing blood products lies in the uncertainty of blood demand and supply, with a trade-off between shortage and wastage, especially in most developing countries. Thus, reliable demand predictions can be imperative in planning voluntary blood donation campaigns and improving blood availability within Ghana hospitals. However, most historical datasets on blood demand in Ghana are predominantly contaminated with missing values and outliers due to improper database management systems. Consequently, time-series prediction can be challenging since data cleaning can affect models' predictive power. Also, machine learning (ML) models' predictive power for backcasting past years' lost data is understudied compared to their forecasting abilities. This study thus aims to compare K-Nearest Neighbour regression (KNN), Generalised Regression Neural Network (GRNN), Neural Network Auto-regressive (NNAR), Multi-Layer Perceptron (MLP), Extreme Learning Machine (ELM) and Long Short-Term Memory (LSTM) models via a rolling-origin strategy, for forecasting and backcasting a blood demand data with missing values and outliers from a government hospital in Ghana. KNN performed well in forecasting blood demand (12.55% error); whereas, ELM achieved the highest backcasting power (19.36% error). Future studies can also employ ML algorithms as a good alternative for backcasting past values of time-series data that are time-reversible.

© 2021 The Author(s). Published by Elsevier B.V. on behalf of International Institute of Forecasters. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

### 1.1. Background of the study

Blood supply chain (BSC) encapsulates all the processes of collecting, testing, processing, storing and distributing blood and its components from donor to recipient patient (Osorio, Brailsford, & Smith, 2015; Stanger, Wilding, Yates, & Cotton, 2012). Blood cannot be manufactured artificially, and supply depends on voluntary human donors who cannot be easily predicted. Given that blood

demand and supply are uncertain, and blood is a limited resource and a perishable product with short shelf life, it is challenging to manage blood inventory. The major challenge in managing blood products lies in the uncertainty of blood demand and supply, with a trade-off between shortage and wastage (Stanger et al., 2012). Blood demand is increasing rapidly among developed countries, such that 10 out of every 100 people in the hospital requires some blood products (The Lancet, 2005). Among developing countries, approximately 100,000 deaths are recorded yearly due to blood shortage and common blood-borne infections from unscreened donated blood (The Lancet, 2005). Thus, reliable demand predictions can be imperative in planning blood donation campaigns and improving

\* Corresponding author.

E-mail address: [twumasic@cardiff.ac.uk](mailto:twumasic@cardiff.ac.uk) (C. Twumasi).

blood availability. Empirical studies on the management of blood products are dominated by the use of operational research methodologies (such as critical path supply chain analysis, optimisation, Markov chains and queueing theory) and other simulation techniques (Katsaliaki & Brailsford, 2007; Kopach, Balcioglu, & Carter, 2008; Van Dijk, Haijema, Van Der Wal, & Sibinga, 2009). Time-series approaches for blood demand forecasting are now gaining popularity in the literature of BSC (Fortsch & Khapalova, 2016; Pereira, 2004). The precision of predictions from time-series models concerning the blood demand has been considered an essential determinant of donor recruitment decision making and inventory control (Shih & Rajendran, 2019). Consequently, a large amount of data is expected to be collected over time to predict blood demand trends and identify underlying patterns such as seasonality through smoothing and time-series decomposition (Pierskalla, 2005; Rajendran & Ravindran, 2019).

Moreover, the extant literature on the blood supply chain has been dominated by studies based on blood centres from developed countries such as the USA (Katsaliaki & Brailsford, 2007; Pereira, 2004), Canada (Kopach et al., 2008), Finland (Ryttilä & Spens, 2006) and Estonia (Alloja, Espenberg, & Kiivet, 2012), among others. However, empirical studies on blood demand forecasting using real-life time-series data in less developed countries are insufficient in the literature. In a less developed country like Ghana, blood products are in high demand for varying health issues. Due to the frequent blood shortages at the major hospitals in Ghana, the family-replacement system has been instituted as the replenishment policy to compel blood donation. However, this replenishment system does not ensure the sufficiency of blood supply. Hence, investigating the requests for blood in developing countries must be approached as a public health issue. Therefore, this paper is based on the premise that a more robust demand predictive method would be a useful planning tool for voluntary blood donation campaigns to supplement supplies and improve blood availability. In Ghana, blood banking and transfusion facilities have poor database management systems predominantly due to lack of computers, computer programs, and training (Nene, Olayemi, & Asamoah-Akuoko, 2015). Thus, most of these health facilities recorded data by hand in books and other documents without the proper filing of patient's medical records, until recent times where the need to collect blood data more efficiently for easy retrieval became a matter of national concern (Nene et al., 2015; Teviu, Aikins, Abdulai, Sackey, Boni, Afari, & Wurapa, 2012). Consequently, most of the past years' data are lost completely, and available data are usually contaminated with missing values and outliers due to genuine recording errors. In such instances, the fidelity of forecasts is greatly affected due to a bias in model parameter estimates and the outliers' carry-over effect on the point forecasts (Chen & Liu, 1993).

Backcasting, a term coined by Robinson (1990), was previously proposed as a planning method in many fields of study, including urban planning and resource management, to investigate future mechanisms from the present by moving backwards in time to determine what policy

measures would be required to reach those future outcomes (Bibri, 2018; Phdungsilp, 2011). In the context of time-series modelling, the backcasting methodology can be adapted to predict lost or unavailable data of past years based on current values by forecasting backwards in time. Thus, backcasting and forecasting are only distinguished based on the direction of predictions. The former retro-polates past values from future values in reverse time, and the latter extrapolates the future values based on historical data. In blood supply chain management, backcasting can be used to ensure the coherence of blood-related data by predicting lost data of past years for short time-series to aid in exploring relevant long-term patterns or trends, which can provide additional insights into inventory management policies, improving data quality for systematic analysis and the effectively assessing the imbalance between blood supply and demand. Classical time-series models cannot capture the non-systematic changes of outliers due to their exogenous effects (López-de Lacalle, 2016). Different types of outliers may exist in time-series data such as additive outlier, innovation outlier, level shift, temporary change and seasonal level shift (Ahmar et al., 2018; Chen & Liu, 1993; López-de Lacalle, 2016). An automatic detection method for the different types of outliers in a given series has been developed (Chen & Liu, 1993), and a more robust data imputation method utilising Kalman smoothing on state-space methods have been developed for univariate time-series data (Durbin & Koopman, 2012; Hinich, 2005; Moritz & Bartz-Beielstein, 2017). Before time-series modelling, data pre-processing is thus vital for model selection, parameter estimation, and predictions.

Machine learning (ML) algorithms for forecasting blood demand series have been recently developing, and their performance against classical time-series models have even been explored in previous studies (Bontempi, Ben Taieb, & Le Borgne, 2013; Papacharalampous, Tyralis, & Koutsoyiannis, 2018; Shih & Rajendran, 2019). Nonetheless, in the context of blood supply management, no known study has explored these ML algorithms' performance in backcasting or reverse forecasting blood demand during instances where past years' data are completely unavailable compared to their forecasting abilities. Moreover, the literature on time-series backcasting is also insufficient relative to forecasting in general. Thus, the predictive power of machine learning models for backcasting past time-series values is also imperative. Moreover, in evaluating the performance of ML algorithms and traditional time-series models, most studies employ the fixed-origin strategy instead of a rolling-origin evaluation with different training and test partitions. However, according to Tashman (2000), the efficiency and reliability of out-of-sample evaluations of time-series models can be improved by adopting rolling-origin strategies. Hence, an out-of-sample rolling strategy is a better approach for generalisability and model comparison. This study thus assesses both forecasts and backcasts from the time-series models using a proposed rolling-origin strategy. For additional information about the essence of rolling-origin assessment approaches in any forecasting studies, see work by Tashman (2000).

In this paper, we compare six different machine learning algorithms (K-Nearest Neighbour regression, Generalised Regression Neural Network, Neural Network Auto-regressive, Multi-Layer Perceptron, Extreme Learning Machines, and Long Short-Term Memory models) for forecasting and backcasting blood demand data with missing values and outliers from a major hospital in Ghana. The standard ARIMA model is used as the baseline model for comparison. We use an out-of-sample rolling-origin strategy with model re-calibration for forecasting and backcasting assessments and statistically compare the prediction error distributions between the time-series models under investigation. Before the time-series modelling, we perform data cleaning by adopting existing state-of-the-art algorithms in two stages: (i) data imputation by Kalman smoothing and (ii) automatic outlier detection and correction. We further justify that the blood demand series under study is time-reversible for backcasting using Teraesvirta's neural network linearity test, Lobato-Velasco normality test of stationary processes, and surrogate data testing.

## 1.2. Paper structure

This study is structured in six main sections. Apart from the background of the study, the paper is organised as follows. First, Section 2 reviews relevant literature on inventory management policies and forecasting methods in BSC; and also presents the main research gap and contribution of this paper. Then, the empirical data, data preprocessing techniques adopted, the proposed backcasting scheme, the time-series models (ARIMA and six ML models) to be used for blood demand predictions, and a summary of the empirical evaluation methods are presented in Section 3. After that, Section 4 presents the main results of the study, and Section 5 discusses the key findings. Finally, we conclude and provide some recommendations for future studies in Section 6.

## 2. Literature

### 2.1. Impact of blood demand forecasting on BSC inventory management policies

The improvement of BSC (outlined in Fig. 1) has become a matter of global concern due to the challenges posed in healthcare delivery. For example, inventory management of blood products is crucial in hospital operations due to their perishable nature and demand uncertainty. Also, the short shelf-life of blood components such as platelets can significantly lead to wastage or expiration if excess platelets are kept. Nonetheless, blood products' inaccessibility when needed may result in loss of lives and other critical conditions. Thus, recent studies have focused on exploring possible solutions for blood demand uncertainty, inventory management and inadequate human resources (Fortsch & Khapalova, 2016). A cost-effective hospital inventory management system should reduce wastage while sustaining the required level of service. Blood shortages in hospitals are predominantly due to the increasing demand and the constrained or irregular

supply (Rajendran & Srinivas, 2020). Additionally, the frequency of demand for blood is regularly rising due to the ageing population and the increasing number of accidents, amongst other reasons (Ali, Auvinen, & Rautonen, 2010). Consequently, blood inventory management should incorporate the challenge of exploring the trade-off between wastage and shortage (Rajendran & Srinivas, 2020). Accurate blood demand forecasting during the development of inventory policies thus aids in lowering costs, reducing blood wastage, and conserving limited resources (Fortsch & Khapalova, 2016).

There exist several generalisable inventory policies for perishable products, and other blood products developed based on inventory management models under demand-supply uncertainties (Dillon, Oliveira, & Abbasi, 2017; Li, Chiang, Down, & Heddle, 2021; Rajendran & Srinivas, 2020; Shokouhifar, Sabbaghi, & Pilevari, 2021). The frequently employed policies in BSC to manage inventory levels in order to bridge the gap between blood shortage and wastage are the periodical policies (Shokouhifar et al., 2021). Other hybrid ordering policies for BSC under demand uncertainty have also been proposed in the literature, and their performances examined (Rajendran & Srinivas, 2020). These inventory policies provide a decision support system for healthcare practitioners to ascertain the best order quantities, given the characteristics of the hospital. In summary, demand forecasting is considered as one of the essential aspects of inventory management, and it can be combined with inventory management policies to obtain optimal strategies even for instances of intermittent blood demands (Ramaekers & Janssens, 2014).

### 2.2. Forecasting methods in blood supply chain

The characteristics of BSC tend to be more complex and dynamic than the traditional supply chain (Jittamai and Boonyanusith (2014). Several quantitative approaches have been adopted by researchers in managing the complexity and uncertainty issues of BSC. Analytical methods, time-series methods, dynamic programming, queuing technique, optimisation, simulation and combination of these methods have been proposed Osorio et al. (2015). Thus, the methods can generally be categorised into three: simulation (eg. artificial neural network and Monte Carlo methods) (Baesler, Nemeth, Martínez, & Bastías, 2014; Beliën & Forcé, 2012; Custer et al., 2004; Khaldi, El, Chihe, Khaldi, El, & Chihe, 2019; Pierskalla, 2005), optimisation (eg. stochastic, linear, dynamic and mixed-integer programming) (De Kort, Janssen, Kortbeek, Jansen, van der Wal, & van Dijk, 2011; Van Dijk et al., 2009; Arvan, Tavakoli-Moghadam, & Abdollahi, 2015; Chaiwutisak, Smith, Wu, Potts, Sakuldamrongpanich, & Pathomsiri, 2016), and time-series methods such as univariate and multivariate moving average models (Critchfield, Connolly, Ziehwein, Olesen, Nelson, & Scott, 1985; Da Veiga, Da Veiga, Catapan, Tortato, & Da Silva, 2014; Fortsch & Khapalova, 2016; Maia, De Carvalho, & Ludermir, 2008; Pereira, 2004; Rasyid & Adhiutama, 2014; Silva-Filho, Carvalho, Cezarino, Salva, & Silviano, 2013).

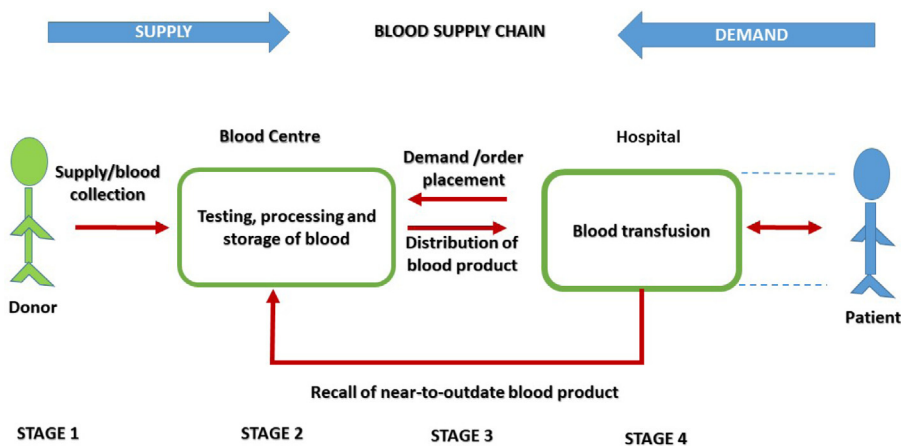


Fig. 1. Blood Supply Chain Processes.

Source: Adapted from Stanger et al. (2012).

### 2.2.1. Literature on time-series methods for blood demand forecasting

Time-series methods have been proposed and applied in forecasting daily, monthly, and yearly blood demand data at hospitals and other health centres. The time-series models for blood demand prediction in the literature have been dominated by the classical univariate time-series models, including autoregressive integrated moving average (ARIMA) models, Holt-Winters' exponential smoothing models and time-series decomposition for both seasonal and non-seasonal data (Pankratz, 2009; Pereira, 2004; Shih & Rajendran, 2019), amongst other methods including logistic regression models (Bosnes, Aldrin, & Heier, 2005). Fortsch and Khapalova (2016) discovered that demand for blood is naturally non-stationary. Therefore, univariate time-series models fitted using the Box-Jenkins methodology based on the integrated moving average (ARIMA) class of models are mostly the optimal choice over the naive, exponential smoothing, simple moving average and other classical forecasting models. However, other studies have discovered limitations in the conventional method in identifying the best specified ARIMA model for forecasting non-seasonal data (Simon, 2007). The conventional approach of best fit does not always guarantee the model's ability to predict future values nor its predictive performance for previous periods due to underlying factors of the data (Simon, 2007).

To efficiently fit a classical ARIMA model, essential data features such as seasonality and trends need to be eliminated to achieve a stationary and non-seasonal series if they exist. Thus, ARIMA and most classical time-series models are not robust unless these models' parametric assumptions are met. One of the most widely employed models is the seasonal ARIMA (SARIMA) model in the presence of seasonality. Both ARIMA and SARIMA models assume that the future values have a linear relationship with current and past series. The ARIMA models assume that its residuals must be independent and identically distributed. The residuals must also have a mean of zero, constant variance and serially uncorrelated (white noise). Thus, for complex non-linear problems,

predictions by either ARIMA or SARIMA may not be adequate. Hence, hybrid models that combine these ARIMA classes of models or other traditional approaches such as Holt-Winter methods and computationally intelligence models such as Artificial Neural Network (ANN) become ideal (Khashei, Bijari, & Hejazi, 2012). Consequently, a different but robust class of time-series models called machine learning algorithms were developed. Few studies have employed machine learning algorithms such as the ANN to predict future demand for blood; their performance against the classical time-series models has been considered (Alajrami, Abu-Nasser, Khalil, Musleh, Barhoom, & Naser, 2019; Darwiche, Feuilloy, Bousaleh, & Schang, 2010; Khaldi, El Afia, Chiheb, & Faizi, 2017). Shih and Rajendran (2019) compared traditional time-series models' performance to ANN and multiple regression via machine learning algorithms on a five-year historical data from Taiwan Blood Services Foundation (TBSF). They discovered that the classical time-series forecasting (predominantly Seasonal Exponential Smoothing Method and ARIMA models) outperformed the machine learning (ML) algorithms. The cause for such under-performance of the ML algorithms is attributed to limited data or small sample size used in training these algorithms; hence, it can lead to biased ML predictions in such instances (Vabalas, Gowen, Poliakoff, & Casson, 2019). Khaldi et al. (2017) found that ANN outperformed ARIMA models for blood demand forecasting. Thus, they concluded that these models could be considered a promising approach to forecasting monthly blood demand. Hence, the classical time-series methods are usually considered baseline models for machine learning algorithms.

However, ML algorithms have been considered in academic literature as suitable alternatives to traditional time-series forecasting models (Makridakis, Spiliotis, & Assimakopoulos, 2018). Nevertheless, minimal evidence about their relative performance (accuracy and computational requirements) is available (Makridakis et al., 2018). Makridakis et al. (2018) compared the post-sample accuracy between Neural Networks (NN), Automated ANN and eight classical models for different forecasting horizons. It



was revealed that the prediction power associated with the machine learning algorithms were relatively better than the traditional methods across all examined forecasts horizons; however, ML computational requirements were more significant than the classical time-series models. Machine learning algorithms for forecasting blood demand data (with linear or non-linear trends) from blood centres in developing countries (including Ghana) are under-studied. Thus, the classical ARIMA model is considered the baseline model for comparing the ML models in this study.

### 2.2.2. Brief review of ML methods for time-series predictions

Machine learning (ML) is a computational intelligence technique that adopts programmed algorithms to analyse input data and learn from it via supervised or unsupervised processes to predict output values within an acceptable range (Wakefield, 2013). ML models have proven to be an excellent alternative to classical statistical models for forecasting and other research problems (such as regression and classification problems) over the last decade. Neural network (NN) commenced the development of ML algorithms. NNs are deep learning methods developed as mathematical models of the brain. They can allow complex non-linear relationships between the response variable and its predictors. Over time, other studies employed the concept of neural networks to develop Decision trees, Random forest, Gradient Boosting Machines, Support vector machines and other ML algorithms for regression-type problems (Alpaydin, 2020; Friedman, Hastie, Tibshirani, et al., 2001). There have been parallel efforts towards empirical validation of existing models, model comparison and development of new ones. The immense importance of these ML developments to modellers provides a wide range of choices and a comprehensive understanding of available models' strengths and weaknesses for different forecasting problems. The ML algorithms for time-series forecasting dominating the literature are Multi-Layer Perceptron (MLP), Bayesian Neural Network (BNN), Radial Basis Functions (RBF), Generalised Regression Neural Network (GRNN), K-Nearest Neighbour regression (KNN), CART regression trees (CART), Support Vector Regression (SVR), Recurrent Neural Network (RNN), Long Short Term-Memory neural network (LSTM), Automated Artificial neural network (AANN) and Gaussian Processes (GP) (Ahmed, Atiya, Gayar, & El-Shishiny, 2010; Makridakis et al., 2018). For an extensive description of these ML models, see works of Ahmed et al. (2010), Alpaydin (2020) and Hastie, Tibshirani, and Friedman (2009). MLP, BNN, GP, KNN and GRNN were found in a large-scale comparative study to be the top five machine learning algorithms based on 18-month one-step-ahead forecasts, among other ML models (Ahmed et al., 2010). However, the differences in their performance could be affected by the choice of predictive error measures (Makridakis et al., 2018) and the type of evaluation strategy employed (Tashman, 2000). The performance of ML models may also vary depending on the historical time-series data and underlying hyper-parameters. Other studies have also explored the predictive performance of ML algorithms such as

Neural Network Auto-Regressive model (NNAR) and Extreme Learning Machine (ELM). ML models for time-series forecasting have significantly evolved over the years and are considered good competitors to the classical models within the forecasting community.

### 2.3. Research gap and contribution

Forecasters, policymakers, time-series users and practitioners usually need long time-series data for model assessment, policy analysis, and investigating underlying trends and patterns to aid decision-making (Caporin & Sartore, 2011). Nevertheless, such long time-series data may not always be obtainable at the expected frequency, with the needed temporal or spatial coverage of the data unavailable in previous years, especially among blood centres in developing countries. The data unavailability may be due to several reasons such as human error, human failure, software corruption, data storage destruction, lack of the required data processors, or data collection only started in some future time, amongst others. Consequently, it is imperative to estimate or predict the lost data of past years for time-series users. The process of predicting data of past years is referred to as backcasting or reverse forecasting. It is possible to backcast or forecast in reverse time for relevant time-series provided the series is strictly stationary and time-reversible (Caporin & Sartore, 2011; Sharifdoust & Mahmoodi, 2013). Unfortunately, to the best of our knowledge, no known study has explored the backcasting power of the existing state-of-the-art ML algorithms for predicting unavailable blood demand data of past years. Hence, this current study attempts to bridge this gap by investigating the forecasting and backcasting power of a few selected ML algorithms (KNN, NNAR, GRNN, MLP, ELM and LSTM) for a short time-series data on blood demand with past lost values using an out-of-sample rolling-origin evaluation strategy for model comparison. The current study can further be expanded by time-series modellers for other general modelling problems with short series contaminated with missing values and outliers, and the findings can help policymakers in the management of the blood supply chain (as previously discussed) as well as provide publicly accessible adaptive programming codes in R (with the help of existing packages) to forecast or backcast any time-series data using the underlying time-series models via the proposed rolling-origin strategy.

The key contributions of this study are:

- i. To demonstrate the application of ML algorithms and a classical time-series model for also backcasting lost data of past years or any time-reversible stationary series.
- ii. To establish that the direction of prediction (forecasting or backcasting) can affect the predictive performance of ML models given time-series data.
- iii. To justify the need for an out-of-sample rolling-origin strategy in comparing existing time-series models' forecasting and backcasting power for short time-series data (with missing values and outliers).

### 3. Materials and methods

As previously discussed, the study primarily compares the predictive performance of six different ML algorithms for forecasting and backcasting blood demand in a major hospital in Ghana based on a short series with missing values and outliers. In addition, this section discusses the data source, data preprocessing procedures, the proposed backcasting scheme based on time-reversibility, the time-series models for predicting blood demand via a rolling-origin strategy, and summarises the empirical evaluation methods. All analyses were carried out using R statistical software version 3.6.3 (R Core Team, 2019). The R scripts developed (for the forecast and backcast schemes via a rolling-origin strategy) and the empirical data used for the study can be found via the GitHub URL link (for reproducibility of results): <https://github.com/twumasiclement/TimeSeries-Forecasting>.

#### 3.1. Sources of data

Monthly data collected from January 2013 to September 2020 on blood demand was obtained from Tema General Hospital (TGH) for this study. TGH has acquired the Enzyme-Linked Immunosorbent Assay (ELISA) machine to screen its blood samples just recently (Ghana Health Service, 2014). Therefore, most data on blood demand before acquiring the ELISA machine remains fragmented and difficult to retrieve (coupled with missing values). From 2014, the hospital started recording the aggregate blood quantity demanded every month at the Health Information Department of the Hospital. Thus, the time-series data for this study is short in length with missing values and outliers (assumed to be due to genuine recording error). The Institutional Review Board of the Ghana Institute of Management and Public Administration Business School (GIMPA) gave ethical approval for the study. The TGH's management also gave the authorisation to access the secondary dataset on aggregated monthly blood demand.

#### 3.2. Data preprocessing

The data cleaning or preprocessing was done in two stages using existing state-of-the-art algorithms. The blood demand series had 14% missing values ( $n = 13$ ) and 11 gaps with an average gap size of 1.182. In the first preprocessing stage, missing value imputation was done by adopting the Kalman Smoothing on a basic structural model (BSM) implemented in the *imputeTS* R package (see Moritz & Bartz-Beielstein, 2017). For the state-space form of the BSM, see work by Durbin and Koopman (2012). Fig. 2 shows the time-series plot of the blood demand series with highlighted missing values and the imputed data at the first data-preprocessing stage. During the second stage of data preprocessing, an automatic outlier detection and adjustment algorithm for identifying additive outliers (AO) and other outlier types (innovation outliers, level shift, temporary change and seasonal level shift) based on the best ARIMA model errors was used (proposed by Chen & Liu, 1993). Additionally, we used the

*tsoutliers* package in R (implemented by López-de Lacalle, 2016) for the automatic ARIMA data correction based on the blood demand series imputed by Kalman smoothing on BSM (from the first stage of data preprocessing). As a result, the automatic ARIMA data correction detected only three additive outliers and no other outlier type based on the best ARIMA model errors (Fig. 3). The final corrected data was used for the subsequent time-series modelling. The Augmented Dickey–Fuller unit root (ADF) test was used to investigate the stationarity of the corrected data. The ADF test revealed that the corrected blood demand series was stationary at lag 0 (Test statistic =  $-4.305$ ,  $p$ -value  $\leq 0.01$ ).

#### 3.3. Time-series forecasting and backcasting models

The principles of time-reversibility and backcasting based on forecasts are discussed (Section 3.3.1). This section also presents the time-series models under investigation (classical non-seasonal ARIMA as baseline model and six ML algorithms) to forecast and backcast blood demand (Sections 3.3.2–3.3.8).

##### 3.3.1. Time reversibility and linearity

**Definition 3.1.** Let  $\{Y_t, t \in \mathbb{Z}\}$  be a strictly stationary series or stochastic process.  $Y_t$  is said to be time-reversible (TR) if, for any positive integer  $n$  and all integers  $t_1, t_2, \dots, t_n$  ( $-\infty < t_1 < t_2 < \dots < t_n < \infty$ ), the vector  $(Y_{t_0}, Y_{t_1}, \dots, Y_{t_n})'$  and  $(Y_{t_n}, Y_{t_{n-1}}, \dots, Y_{t_0})'$  have the same joint distribution; such that

$$(Y_{t_0}, Y_{t_1}, \dots, Y_{t_n})' \stackrel{D}{\sim} (Y_{t_n}, Y_{t_{n-1}}, \dots, Y_{t_0})', \quad (1)$$

where  $\stackrel{D}{\sim}$  represents equal in distribution (Sharifdoust & Mahmoodi, 2013).

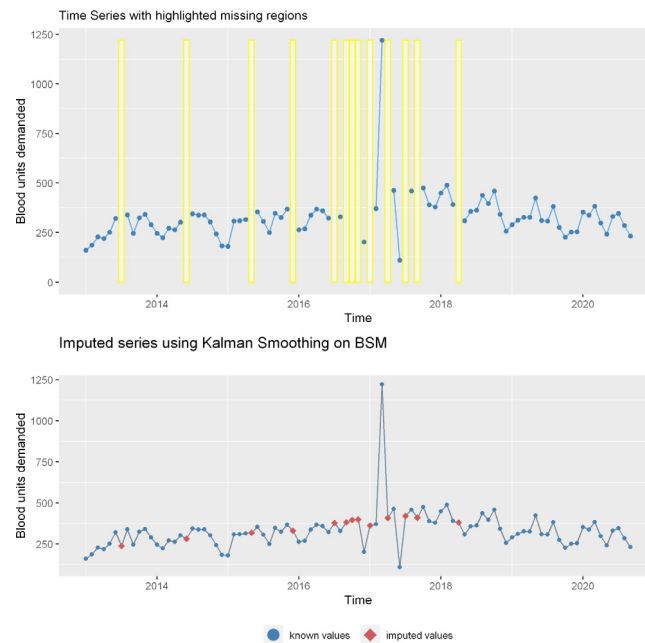
**Definition 3.2.** The stochastic process  $\{Y_t, t \in \mathbb{Z}\}$  is said to be a linear process if it can be represented as

$$Y_t = \sum_{k \in \mathbb{Z}} \varphi_k Z_{t-k} \quad (2)$$

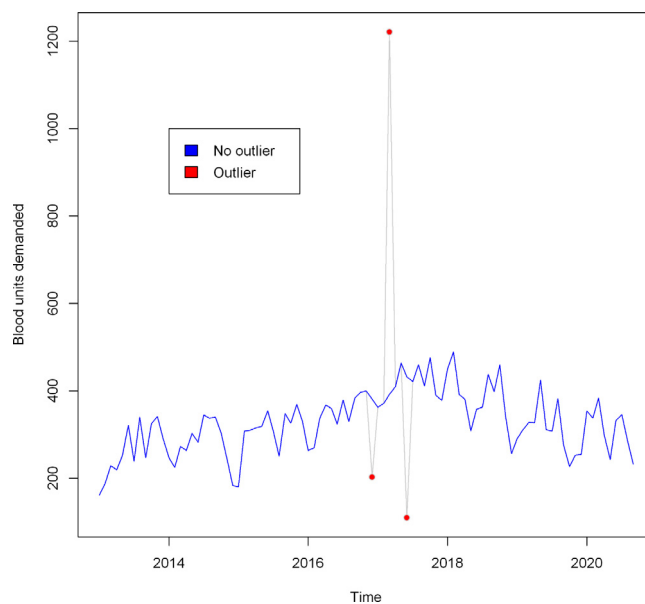
where  $\{Z_t, t \in \mathbb{Z}\}$  is a sequence of nondegenerate independent and identically distribution random variables with mean 0 and constant variance  $\sigma^2$ ; and  $\{\varphi_k\}$  is a sequence of constants such that  $0 < \sum_{k \in \mathbb{Z}} \varphi_k^2 < \infty$  (Sharifdoust & Mahmoodi, 2013).

**Lemma 1.** Let  $\{Y_t, t \in \mathbb{Z}\}$  be a strictly stationary linear process such that  $Y_t = \sum_{k \in \mathbb{Z}} \varphi_k Z_{t-k}$ ; then  $\{Y_t, t \in \mathbb{Z}\}$  is time-reversible if and only if it is a Gaussian process or for  $t_0 \in \mathbb{Z}$  and  $\rho = \pm 1$ ,  $\varphi_t = \rho \varphi_{t_0-t}$  and  $Z_t \stackrel{D}{\sim} \rho Z_t$  (Sharifdoust & Mahmoodi, 2013).

**Remark.** Hence, to show that a strictly stationary (blood demand) series is time-reversible and thus backcasting with the ARIMA and ML models is possible, we need to establish that the stationary series follows a Gaussian process, comes from a linear stochastic process, and there is linearity in the mean. Once this assumption of time reversibility holds, backcasting with the time-series



**Fig. 2.** Units of blood demanded from January 2013 to September 2020 with highlighted missing regions (top) and imputed series using Kalman smoothing on BSM (bottom).



**Fig. 3.** Corrected blood demand series by automatic ARIMA data correction algorithm.

models is done by forecasting with the reversed series in time. Lobato-Velasco normality test of stationary process (LV test) was used to test the Gaussianity of the stationary blood demand series using the *nortsTest* package in R (Asael Alonzo Matamoros & Nieto-Reyes, 2020). In addition, Teraesvirta's neural network linearity test, which tests whether there is linearity in mean given the series (Constantino, Garcia, & Sawitzki, 2020; Teräsvirta, Lin, & Granger, 1993) and the surrogate data testing,

which tests whether the series is a Gaussian linear process (Constantino et al., 2020; Schreiber & Schmitz, 2000) were employed at  $\alpha = 5\%$ . For the Surrogate data testing, the time symmetry statistic ( $T$ ), which measures the asymmetry of stationary time-series ( $Y_t$  of length  $n$ ) under time-reversibility, is computed as (adapted from Kantz and Schreiber (2004)):

$$T = \frac{1}{n-1} \left[ \sum_{t=1}^{n-1} Y_t Y_{t+1}^2 - \sum_{t=1}^{n-1} Y_t^2 Y_{t+1} \right] \quad (3)$$

Surrogate data of size  $\frac{2K}{\alpha} - 1$ , for  $K \geq 1$  is generated using a phase randomization procedure (based on Fast Fourier Transform of the  $Y_t$ ) for a two-sided test (Constantino et al., 2020; Raeth & Monetti, 2009). The null hypothesis (the series is a Gaussian linear process) for a two-sided Surrogate test is rejected if the test statistic for the original series  $Y_t$  is significantly different from test statistics for all the  $\frac{2K}{\alpha} - 1$  generated surrogate datasets.

The LV normality test revealed that the stationary blood demand series was normally distributed (Test statistic = 0.078, df = 2, p-value = 0.962). Teraesvirta's neural network linearity test revealed that the corrected blood demand series ( $\chi^2 = 0.647$ , df = 2, p-value = 0.723) is linear in its mean. The symmetrical distribution of the time-reversibility statistic (based on the generated surrogate and observed series) and the normally distributed demand series are shown in Fig. 4. The time-reversibility statistic for the corrected blood demand series is not significantly different from test statistics of all generated surrogate data (p-value > 0.05). Therefore, the blood demand series comes from a Gaussian linear process and is thus time-reversible. Thus, blood demand backcasting with time-series models is possible since the data is confirmed to be time-reversible. Moreover, according to Sharifdoust and Mahmoodi (2013), if a stationary linear model is assumed, then a test for time reversibility is usually equivalent to a test for Gaussianity of the stationary series. In conclusion, the time-reversibility assumption is a necessary condition for the proposed backcasting scheme using the time-series models since backcasts are obtained by reversing forecasts backwards in time (as previously justified by Sharifdoust & Mahmoodi, 2013). This is because, if the stationary series is time-reversible, the joint probabilities of the forward and reverse state sequences from the time-series models are the same for all sets of time increments.

**Proposition 3.1.** Let  $Y_t = (Y_m, Y_{m+1}, \dots, Y_{m+n})'$  for  $t \in \{m, m+1, \dots, m+n\}$  and  $m, n \in \mathbb{Z}$  (where  $0 \leq m < n < \infty$ ), be the current series of a stationary data with frequency  $f$  and time-reversible. Suppose  $F_t^h = (\hat{Y}_{m+n+1}, \hat{Y}_{m+n+2}, \dots, \hat{Y}_{m+n+h})'$  are the one-step-ahead forward prediction for  $h$  horizons based on a fitted time-series model; then the backcasts of past values for  $h$  ( $h \geq 1$ ) horizons at time  $t \in \{m-h, m-h+1, \dots, m-1\}$  are given as:  $B_t^h = (\hat{Y}_{m-h}, \hat{Y}_{m-h+1}, \dots, \hat{Y}_{m-1})'$  such that  $B_i^h = F_j^h$  at past time index  $i = m-h, m-h+1, \dots, m-1$  and reversed time index  $j = m+n+h, m+n+h-1, \dots, m+n+1$  based on the frequency ( $f$ ) of the series.

The pseudo-code summarising the proposed backcasting scheme in this study (in accordance with Proposition 3.1) is given by Algorithm 1.

### 3.3.2. Non-seasonal ARIMA model

Let suppose  $Y_t$  is the observed series at time  $t$ , then the full ARIMA( $p, d, q$ ) model which integrates both AR( $p$ ) and MA( $q$ ) models is given as (Hyndman & Athanasopoulos, 2018):

$$\Delta Y_t = h_0 + \sum_{i=1}^p \alpha_i \Delta Y_{t-i} + \epsilon_t + \sum_{j=1}^q \theta_j \epsilon_{t-j} \quad (4)$$

### Algorithm 1: Pseudo-code for the proposed backcasting scheme using time-series models

---

**Input:** Current series of a stationary data  
 $Y_t = (Y_m, Y_{m+1}, \dots, Y_{m+n})'$  for  
 $t \in \{m, m+1, \dots, m+n\}$  and  $m, n \in \mathbb{Z}$   
 (where  $0 \leq m < n < \infty$ ), with frequency  $f$ .

**Output:** Backcasts of past values ( $B_t^h$ ) for  $h$  ( $h \geq 1$ ) horizons at time  
 $t \in \{m-h, m-h+1, \dots, m-1\}$  such  
 that:  $B_t^h = (\hat{Y}_{m-h}, \hat{Y}_{m-h+1}, \dots, \hat{Y}_{m-1})'$ , and  
 $B_i^h = F_j^h$  at past time index  
 $i = m-h, m-h+1, \dots, m-1$  and  
 reversed time index  
 $j = m+n+h, m+n+h-1, \dots, m+n+1$   
 based on the frequency ( $f$ ) of the series;  
 where the forward prediction  
 $F_t^h = (\hat{Y}_{m+n+1}, \hat{Y}_{m+n+2}, \dots, \hat{Y}_{m+n+h})'$ .

- 1 Test for stationarity of the original series ( $Y_t$ ) or transform series into a stationary form by differencing if necessary.
- 2 Test the Gaussianity of the stationary series using appropriate test.
- 3 Determine if the stationary series is time-reversible by justifying whether the stationary series is a linear Gaussian stochastic process.
- 4 **if stationary series is time-reversible then**
- 5     Fit the required time-series model based on the training set or current series.
- 6     Generate a one-step-ahead forecast for  $h \geq 1$  horizons from the fitted model.
- 7     Reverse the forward predictions or forecasts in time (based on the frequency of the series) for the required backcasts or predicted values of past years.
- 8     Assess the predictive performance of the model based on the validation set and appropriate error measure (preferably via a rolling-origin evaluation strategy for a short series).
- 9 **else**
- 10    backcasting is not possible with the time-series model given a time-irreversible series.
- 11 **end**

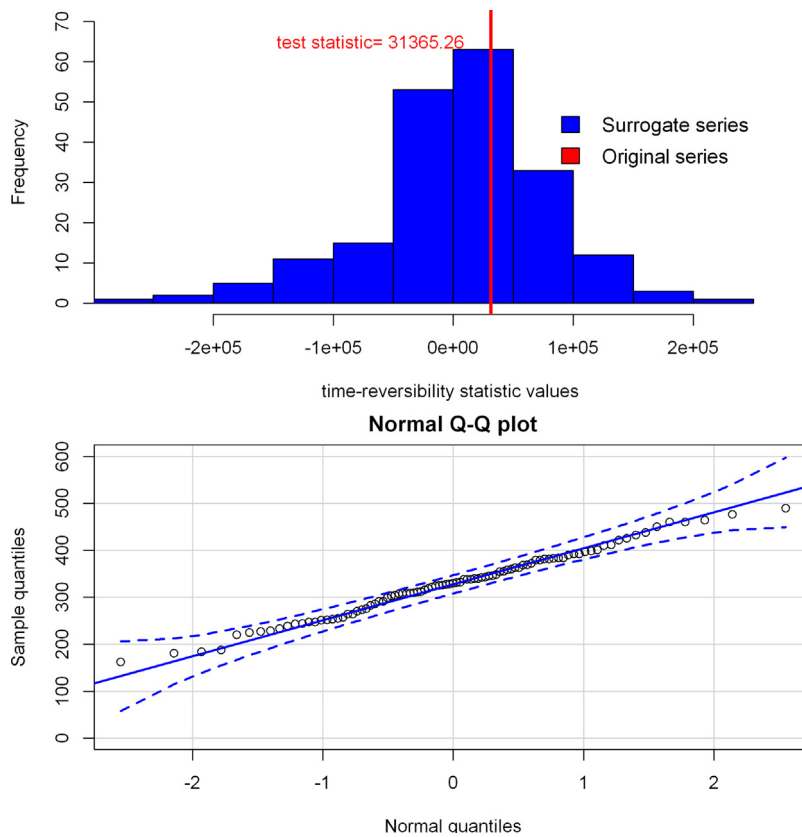
---

where  $\Delta Y_t$  is the differenced series (at differencing order  $d$  denoting the minimum non-negative order such that the series  $Y_t$  is stationary),  $h_0 \in \mathbb{R}$  is a constant and  $\epsilon_t$  is the random error at time  $t$ ; whereas  $p$  and  $q$  are the autoregressive and moving average non-negative orders respectively, and  $\alpha_i \in \mathbb{R}$  for  $i = 1, 2, \dots, p$  and  $\theta_j \in \mathbb{R}$  for  $j = 1, 2, \dots, q$  are the regression coefficients. The *auto.arima* function in R was used to fit an automatic non-seasonal ARIMA model (Hyndman et al., 2020). A function in R was developed to implement the automatic ARIMA via a rolling-origin strategy for model comparison (for both forecast and backcast schemes).

### 3.3.3. K-nearest neighbour regression model

K-Nearest Neighbour regression (KNN) is a non-parametric ML model that makes predictions based





**Fig. 4.** Plot of time-reversibility statistic for the blood demand series and 199 generated surrogate data (top), and Normal Q-Q plot of the blood demand series (bottom).

on the target output of the  $K$  nearest neighbours of the given query point (Ahmed et al., 2010). Now, suppose the  $i$ th training instance or data point consists of a vector of  $v$  features:  $f = (f_1^i, f_2^i, \dots, f_v^i)$  which describes the instance. Given a new instance with known features:  $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_v)$ , but unknown target vector, these new features are used to find its  $K$  most identical training instances from their vectors of features and the Euclidean distance  $d(f, \gamma)$  (a similarity metric) such that:

$$d(f, \gamma) = \sqrt{\sum_{j=1}^v (f_j^i - \gamma_j)^2} \quad (5)$$

Thus, the closest  $K$  training data points are selected, and a prediction is obtained based on the average of the target output values for these  $K$  points (Martínez, Frías, Charte, & Rivera, 2019a; Martínez, Frías, Pérez, & Rivera, 2019b). That is, let assume that the targets  $(y_1, y_2, \dots, y_K)$  are the  $K$  nearest neighbours for a new instance  $x$ , then the prediction ( $\hat{y}$ ) is given by

$$\hat{y} = \frac{1}{K} \sum_{j=1}^K y_j \quad (6)$$

Larger values of the hyperparameter  $K$  lead to a smoother fit at the cost of a higher bias and vice versa for smaller values of  $K$ ; and thus controls the bias–variance

trade-off (Taieb, Bontempi, Atiya, & Sorjamaa, 2012). There are several methods for determining the value of  $K$ , including but not limited to: (i) using some heuristic or rule-of-thumb technique that recommends setting  $K$  to be the square root of the number of training examples (Martínez et al., 2019a), (ii) using a cross-validation or optimisation techniques that estimate the optimal  $K$  value by minimising a forecast error measure such as MAPE (Hyndman & Koehler, 2006), and (iii) using local learning techniques to adaptively set  $K$  by minimising the Leave-One-Out (LOO) error statistic (described by Taieb et al., 2012) or, preferably, the Prediction Sum of Square (PRESS) statistic, which produces forecasts with the most similar stochastic characteristics to the training samples (Allen, 1974; Bontempi, Birattari, & Bersini, 1999). In this study, KNN was fitted using the *tsfknn* package in R (Martínez et al., 2019a). Due to ease of computation and the *tsfknn* package being used, the optimal values of the hyperparameter  $K$  were iteratively tuned from a range of possible values (from 1 to the maximum number of training samples) via a cross-validation procedure which minimises the MAPE statistic (defined in Eq. (20)) based on a multiple-step ahead strategy known as the Multiple-Input-Multiple-Output (MIMO) (described in Bontempi & Taieb, 2011; Taieb et al., 2012). In addition, we developed a function in R to implement KNN via a rolling-origin strategy for model comparison (for both forecast and backcast schemes) by adapting the *tsfknn* R package.

### 3.3.4. Generalised Regression Neural Network model

Generalised Regression Neural Network (GRNN) is a type of ANN model which is more robust for non-linear fitting and originally developed to complement the ARIMA model (BuHamra, Smaoui, & Gabr, 2003; Leung, Chen, & Daouk, 2000). GRNN is a type of Radial Basis Function network (RBF) characterised by a fast single-pass learning, and consists of hidden layers of RBF neurons. Now, given a training set with  $v$  training patterns  $(x_1, x_2, \dots, x_v)$  and its corresponding targets  $(y_1, y_2, \dots, y_v)$ , GRNN makes predictions for an input  $x$  based on the weighted average of target outputs of the training data points in the neighbourhood of  $x$ , using some kernel function (Ahmed et al., 2010). Thus, the prediction ( $\hat{y}$ ) for data point  $x$  is estimated such that:

$$\hat{y} = \sum_{j=1}^v \omega_j y_j \quad (7)$$

where the weights  $w_j$  produced by the hidden layers are given by

$$\omega_j = \frac{\text{Kerf}\left(\frac{\|x - x_j\|}{\rho}\right)}{\sum_{j'=1}^v \text{Kerf}\left(\frac{\|x - x_{j'}\|}{\rho}\right)} \quad (8)$$

where  $y_j$  is the target output for training data point  $x_j$ ,  $\text{Kerf}$  is a kernel function with bandwidth  $\rho$  and  $\|\cdot\|$  is the Euclidean norm. The Gaussian kernel in the form  $\text{Kerf}(\mathbf{u}) = e^{-\mathbf{u}^2/2}/\sqrt{2\pi}$  is typically used. The normalised weights  $\omega$  denote the training pattern contributions to the final output. The hyperparameter of GRNN, the bandwidth  $\rho$ , determines the smoothness of the fit; such that very large  $\rho$  result in prediction close to the mean of training targets (with similar weights), and small  $\rho$  assigns significant weights to training targets closer to the input vector. The GRNN model was fitted, the smoothing parameter  $\rho$  automatically tuned via optimisation using *tsfgrnn* package in R (Francisco Martinez, 2019). Also, a function was developed in R to implement GRNN via a rolling-origin strategy for model comparison (for both forecast and backcast schemes).

### 3.3.5. Neural Network Auto-Regressive model

Neural Network Auto-Regressive model (NNAR) is also a type of artificial neural network (ANN) which allows the modelling of complex non-linear relationships between input and output variables but uses lagged values of the current series as inputs for model fitting (Thoplan, 2014). For a seasonal data with a frequency of  $s$ , the NNAR seasonal model is represented as  $\text{NNAR}(p, P, k)_s$ ; where  $p$ ,  $P$  and  $k$  model parameters denote the trend auto-regressive order, seasonal trend autoregressive order and the number of nodes in the hidden layer, respectively (Hyndman & Athanasopoulos, 2018). The non-seasonal model is denoted by  $\text{NNAR}(p, k)$ . More generally,  $\text{NNAR}(p, P, k)_s$  has inputs  $(y_{t-1}, y_{t-2}, \dots, y_{t-p}, y_{t-s}, y_{t-2s}, \dots, y_{t-ps})$  and  $k$  neurons or nodes in the hidden layer. Eliminating the hidden layer of the model (i.e. setting  $k = 0$ ),  $\text{NNAR}(p, P, 0)_s$  is analogous to the seasonal ARIMA  $(p, 0, 0)(P, 0, 0)_s$  model, and  $\text{NNAR}(p, 0)$  is similar to  $\text{AR}(p)$  model but with non-linear functions (Faraway & Chatfield, 1998). The

NNAR model is a feedforward neural network, where each layer of neurons or nodes receives inputs or training samples from the previous layers. Consequently, the outputs of the nodes in one layer are inputs to the next layer. Now, let suppose  $\mathbf{y}_{t-1} = (y_{t-1}, y_{t-2}, \dots, y_{t-p})$  is an input vector for a single-hidden-layer  $\text{NNAR}(p, k)$  model with  $k$  hidden nodes at time  $t - 1$ ; then, the (non-linear) relationship between the model output ( $y_t$ ) at time  $t$  and inputs ( $\mathbf{y}_{t-1}$ ) at time  $t - 1$  has the following mathematical representation:

$$y_t = \varphi \left( b_i + \sum_{j=1}^p w_{i,j} \cdot y_{t-j} \right) + \epsilon_t \quad \text{for } i = 1, 2, \dots, k \quad (9)$$

where  $b_i$  is the hidden layer bias, and  $w_{i,j}$  is the weight corresponding to the input vector  $\mathbf{y}_{t-1}$  for the  $j$ th ( $j = 1, 2, \dots, p$ ) input in hidden node  $i$ ;  $\epsilon_t$  is the series error (assumed to be homoscedastic and possibly normally distributed), and  $\varphi$  is a logistic sigmoid function in the form

$$\varphi(\mathbf{u}) = \frac{1}{1 + e^{-\mathbf{u}}} \quad (10)$$

The hidden layer bias and weights are numerically estimated from the data (using a backpropagation algorithm and a cost function). Thus, the weights have no closed-form and meaningful interpretation. In this study, a non-seasonal  $\text{NNAR}(p, k)$  model with a single-hidden layer was fitted using the *forecast* package in R (Hyndman et al., 2020). A novel function in R was also developed to implement NNAR via a rolling-origin strategy for model comparison (for both forecast and backcast schemes). The best average number of hidden nodes was determined at different possible values.

### 3.3.6. Multi-layer perceptron model

A multilayer perceptron (MLP), just like NNAR model, is also a feedforward ANN class that utilizes supervised learning techniques via backpropagation for non-linear prediction of a stationary time-series (Rosenblatt, 1961; Rumelhart, Hinton, & Williams, 1985). MLP consists of at least three layers of nodes: an input layer, a hidden layer and an output layer, including hidden node ( $i$ ), weights ( $w$ ) and a transfer function  $\varphi$  (a logistic sigmoid function given by Eq. (10)). Suppose  $x$  is an input vector or training samples of length  $\vartheta$ ; then, the prediction for the network output ( $\hat{\mathbf{y}}_i$ ) for hidden node  $i$  is given as (related to NNAR output prediction):

$$\mathbf{y}_i = \varphi \left( b_i + \sum_{j=1}^{\vartheta} w_{i,j} x_j \right) \quad \text{for } i = 1, 2, \dots, k \quad (11)$$

where  $b_i$  is the hidden layer bias,  $w_{i,j}$  is the weight corresponding to  $j$ th input  $x_j$  ( $j = 1, \dots, \vartheta$ ) for hidden node  $i$ , and  $k$  is the number of hidden nodes. The hidden layer bias and weights can be adjusted based on corrections that minimize the error in the entire output via gradient descent. A major drawback of MLP is its high computational time for training. Hence, the model is trained in

three stages: a forward pass, calculation of error or loss and a backward pass. An automated MLP was fitted using the *nnfor* package in R (Kourentzes, 2017). A 5-fold cross-validation was used to automatically choose the number of hidden nodes and the differencing order of the training samples. A function was created in R to implement MLP via a rolling-origin strategy for model comparison (for both forecast and backcast schemes).

### 3.3.7. Extreme Learning Machine

Extreme Learning Machine (ELM) is also a hidden layer feedforward neural network that does not require gradient-based backpropagation optimisation but Moore–Penrose generalised inverse to set the weights connecting the inputs to the hidden layer and hidden layer biases (unlike other ML algorithms like NNAR or MLP); hence, generally faster in computational time (Huang, Zhu, & Siew, 2006; Jayaweera & Aziz, 2018). The mathematical model for a single-layer ELM with input vector or training samples  $x$  and output ( $y$ ) is given as

$$y_j = \sum_{i=1}^k \beta_i g_i(x) = \sum_{i=1}^k \beta_i g(b_i + \omega_i x_j) \quad \text{for } j = 1, 2, \dots, \vartheta \quad (12)$$

where  $k$  is number of hidden nodes,  $\vartheta$  is the length of the input vector,  $b$  is the hidden layer bias vector,  $\omega$  is the weight vector between input and hidden layer,  $\beta$  is weight vector between the hidden layer and output, and  $g$  is a linear activation function (where  $g$  can be a penalised linear regression or the traditional linear regression). The main hyperparameter that requires tuning is the number of hidden nodes. An automated ELM was fitted using the *nnfor* package in R (Kourentzes, 2017). A 5-fold cross-validation was used to automatically select the number of hidden nodes and the differencing order of the training samples. In addition, the classical Lasso regression was used as the linear activation function ( $g$ ) in estimating the output layer weights. A function was also developed in R to implement ELM via a rolling-origin strategy for model comparison (for both forecast and backcast schemes).

### 3.3.8. Long Short-Term Memory network

Long Short-Term Memory network (LSTM) is a particular type of Recurrent Neural Network (RNN) for dealing with the vanishing gradient problem of the classical RNN or capable of learning long-term dependencies (Gopika, Sowmya, Gopalakrishnan, & Soman, 2020). LSTM comprises memory blocks (or cells) connected through layers. The information in the cells is contained in cell state  $c_t$  and hidden state  $y_t$  at time  $t$  via mechanisms known as gates based on two sigmoid activation functions: hyperbolic tangent function ( $\tanh$ ) and a logistic sigmoid function. LSTM requires the series to be in a supervised learning mode with target ( $y$ ) and predictor ( $x$ ) variables. Consequently, the series is transformed by lagging (at lag  $\phi$ ) such that the lagged value at time  $t - \phi$  is used as the input and lagged value at time  $t$  considered as the target, for a  $\phi$ -step lagged dataset. Additionally, LSTM has three gates: (i) input gate (receives new information and

the prior predictions as inputs), (ii) forget gate (eliminates the information that is no longer relevant in the cell state), and (iii) output gate (makes a selection based on the new information and the previous predictions). Thus, the final output is obtained through three steps or mechanisms described as follows:

Step 1 (input gate  $i_t$ ): Let  $x_t$  (current input at time  $t$ ) and  $y_{t-1}$  (previous hidden state at time  $t - 1$ ) be input information. The logistic sigmoid layer creates an update filter such that

$$i_t = \varphi(W_i x_t + R_i y_{t-1} + b_i) \quad (13)$$

where  $W$  is the input weight vector,  $R$  is the recurrent weight vector,  $b$  is the hidden layer bias vector, and  $\varphi$  is the logistic sigmoid function (given by Eq. (10)).

Step 2 (forget gate  $f_t$ ): This gate eliminates the information that is no longer relevant in the cell state ( $c_t$ ) given weights ( $W_f$ ,  $R_f$  and  $b_f$ ) such that

$$f_t = \varphi(W_f x_t + R_f y_{t-1} + b_f) \quad (14)$$

and the  $\tanh$  activation layer creates a vector of potential candidates ( $z_t$ ) given weights ( $W_z$ ,  $R_z$  and  $b_z$ ) such that

$$z_t = \tanh(W_z x_t + R_z y_{t-1} + b_z). \quad (15)$$

A sigmoid layer then creates an update filter ( $u_t$ ) given weights ( $W_u$ ,  $R_u$  and  $b_u$ ) such that

$$u_t = \varphi(W_u x_t + R_u y_{t-1} + b_u) \quad (16)$$

and the previous cell state  $c_{t-1}$  is updated such that

$$c_t = f_t \otimes c_{t-1} + u_t \otimes z_t \quad (17)$$

Step 3 (output gate  $o_t$ ): The sigmoid layer filters the cell state ( $c_t$ ) for the output ( $o_t$ ) given weights ( $W_o$ ,  $R_o$  and  $b_o$ ) such that

$$o_t = \varphi(W_o x_t + R_o y_{t-1} + b_o). \quad (18)$$

Finally, an element-wise product ( $\otimes$ ) of the scaled cell state ( $c_t$ ) and the filtered output ( $o_t$ ) gives the new hidden state  $y_t$  for the next cell ( $c_{t+1}$ ) such that

$$y_t = o_t \otimes \tanh(c_t). \quad (19)$$

With the help of the open-source R software libraries: *Keras* and *TensorFlow* (Arnold, 2017), a function was created in R to implement LSTM based on the empirical time-series data (lagged at differencing order  $\phi = 1$  and normalised afterwards) via a rolling-origin strategy for model comparison (for both forecast and backcast schemes).

Unlike the other aforementioned R packages for machine learning forecasting (discussed in Sections 3.3.3–3.3.7), the *Keras* and *TensorFlow* libraries do not directly return a forecast object in R when implementing LSTM. Thus, users must take care when fitting LSTM due to numerous modelling steps from time-series differencing, data normalisation, data transformation into 3-dimensional arrays, inverse scaling after obtaining predictions from the compiled model, converting predictions into a forecast object in R, and finally estimating prediction error (either via the rolling-origin forecast or backcast strategy).

### 3.4. Empirical evaluation

The forecasting and backcasting power of six ML algorithms (KNN, NNAR, GRNN, MLP, ELM and LSTM) based on the corrected blood demand data with short length are primarily examined. The non-seasonal ARIMA model was considered as a baseline model for comparison with the ML algorithms. We further investigated whether there exist seasonality in the stationary series using the Webel-Ollech overall seasonality test (implemented in [Ollech & Webel, 2020](#)). The Webel-Ollech test found no seasonality in the corrected blood demand series ( $p$ -value  $> 0.05$ ). This finding is also confirmed in the Auto-correlation Function (ACF) and seasonal plots ([Fig. 5](#)). Also, the stationary blood demand series was found to be time-reversible. Thus, blood demand backcasting (presented by Algorithm 1) with the time-series models was possible under the time-reversibility assumption.

An out-of-sample rolling-origin strategy with model recalibration is used to assess the forecasting and backcasting power of the time-series models under study at different lead times. Thus, the corrected blood demand series was split into different training and test partitions (determined by the maximal and minimal prediction horizons or lead times) according to the following proposed series splitting rule:

Let  $H$  be the maximal prediction horizon (set at  $H = 18$  months), and assume  $\nu$  is the minimal prediction horizon (set at  $\nu = 2$  months). Suppose  $N$  is the total length of the observed series. For forecasting and backcasting schemes, different training sets for model fitting are chosen with increasing lengths  $N - H, N - H - 1, \dots, N - \nu$ ; whereas the test set for cross-validation is chosen at decreasing lead times  $H, H - 1, \dots, \nu$ . Per this proposed rolling-origin strategy, the expected number of different sets of predictions (forecasts or backcasts) or the number of times an individual model is re-calibrated is  $\eta = H - \nu + 1$  (i.e.  $\eta = 18 - 2 + 1 = 17$  prediction sets). Hence, the overall total number of either forecasts or backcasts for this rolling-origin strategy with a maximal prediction horizon of  $H$  is given  $\frac{\eta(H+\nu)}{2}$  (based on the sum of  $\eta$  lead times as arithmetic sequences).

For each set of the 17 different predictions ( $\eta = 17$ ), their respective test set was compared with its corresponding predictions by estimating the mean absolute percentage error (MAPE) given by Eq. (20) at each lead time. A pooled average error statistic (median MAPE denoted by MdMAPE) given by Eq. (21) (recommended by [Tashman & Kruk, 1996](#)) was finally computed for model comparison. Thus, MdMAPEs at the different lead times were obtained for each time-series model during forecasting and backcasting assessments, respectively, resulting in an error distribution. The prediction error distributions of the time-series models were further compared, and the significant differences between the time-series models' error distributions were determined using the standard Kruskal–Wallis (KW) median test and Bonferroni Dunn's test of multiple comparisons at 5% alpha level. For each test set ( $Y_T$ ) at lead time  $T = H, H - 1, \dots, \nu$ , let assume the test series is given by  $Y_T = (Y_1, Y_2, \dots, Y_{n_T})$  and its corresponding predicted values  $F_T = (F_1, F_2, \dots, F_{n_T})$ ,

then the MAPE at lead time  $T$  and the median MAPE (MdMAPE) are computed such that:

$$\text{MAPE}_T = \frac{1}{n_T} \sum_{i=1}^{n_T} \left| \frac{Y_i - F_i}{Y_i} \right| \times 100 \quad \text{for } i = 1, 2, \dots, n_T \quad (20)$$

and

$$\text{MdMAPE} = \text{median}(\text{MAPE}_1, \text{MAPE}_2, \dots, \text{MAPE}_{n_T}) \quad (21)$$

where  $n_T$  is the test series length at lead time  $T$ .

## 4. Results

### 4.1. Blood demand forecasts and backcasts from the ARIMA and ML models

The non-seasonal ARIMA model was considered a baseline model for comparison with the ML models for forecasting and backcasting the units of blood demanded via the proposed 18-month rolling-origin strategy with 17 different sets of predictions; and thus, 17 different fitted ARIMA models. [Fig. 6](#) is a plot of the blood demand forecasts and backcasts at the maximal 18-month prediction horizon.

Moreover, the six ML algorithms (KNN, GRNN, NNAR, MLP, ELM and LSTM) were also fitted to forecast and backcast the units of blood demanded at the Tema General Hospital via the rolling-origin strategy. Comparative plots of the blood demand forecasts and backcasts for the fitted ML models at the maximal 18-month prediction horizon are given by [Figs. 7 and 8](#), respectively.

### 4.2. Overall predictive performance of the time-series models

The predictive power of the time-series models for forecasting and backcasting the short blood demand series were assessed at different prediction origins based on the rolling-origin evaluation (using Eqs. (20) and (21)). [Fig. 9](#) shows that there was some degree of variability in the prediction error at the different forecast and backcast origins for any given time-series model. For example, KNN generally had the highest forecasting power at forecast origins from May 2019 to July 2020; whereas, ELM achieved the best backcasting power at backcast origins from July 2013 to June 2014. The prediction error distribution between ARIMA and the six ML models were also compared as displayed in [Fig. 10](#). It was found from the Kruskal–Wallis median test that there is a significant difference in the median prediction error of the time-series models for both forecasts ( $\chi^2 = 39.058$ ,  $df = 6$ ,  $p$ -value  $< 0.001$ ) and backcasts ( $\chi^2 = 43.081$ ,  $df = 6$ ,  $p$ -value  $< 0.001$ ) at 5% alpha level. In addition, the Bonferroni Dunn's test for pairwise comparison of the models' forecast error distributions ([Table 1](#)) revealed that KNN's forecast error was significantly different (and lower) than the forecast errors of the ARIMA model and the other machine learning algorithms (NNAR, GRNN, MLP, ELM and LSTM models).

Also, pairwise comparison of the models' backcast error distributions ([Table 2](#)) showed that, except for the



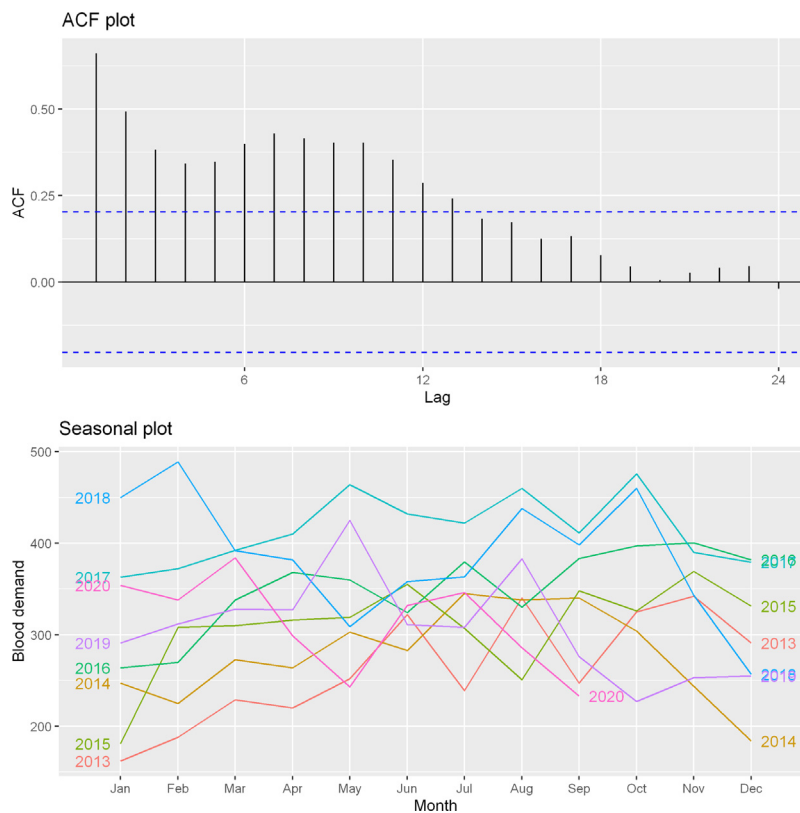


Fig. 5. Auto-correlation function (top) and seasonal (bottom) plots based on the corrected series.

KNN model, the backcast errors of ELM were significantly different (and lower) than the backcast errors of the other time-series models; whereas, backcast errors of NNAR and KNN were significantly different at 5% alpha level. We can therefore conclude from Table 3 that the best forecasting model was KNN (12.55% error), followed by NNAR (17.63% error) and ARIMA (18.40% error). On the other hand, ELM achieved the highest backcasting power (19.36% error), followed by KNN (25.94% error) and LSTM (28.21% error). Nevertheless, there was no significant difference between the forecasting power of the ML algorithms and the ARIMA model, except the KNN model (Table 1). However, ELM significantly outperformed the ARIMA model in backcasting blood demand based on the short time-series data. The high variability in the models' predictive power at the different training and test partitions of the data suggests the need to adopt the rolling-origin strategies over the fixed-origin strategy for effective and robust model comparisons (Figs. 9 and 10). From Table 3, it can also be inferred that all the time-series models under investigation performed well in forecasting blood demand than backcasting based on the given empirical data.

## 5. Discussion

### 5.1. Theoretical implications of the study

This study mainly investigated the predictive performance of ML algorithms for forecasting and backcasting

Table 1

P-values of the Bonferroni Dunn's test for pairwise comparison of models' forecast error distributions.

Models (i, j)	ARIMA	KNN	NNAR	GRNN	MLP	ELM
KNN	0.001					
NNAR	1.000	0.002				
GRNN	1.000	0.000	1.000			
MLP	0.602	0.000	1.000	1.000		
ELM	1.000	0.000	1.000	1.000	1.000	
LSTM	1.000	0.000	1.000	1.000	1.000	1.000

\*  $p$ -value < 0.05 in  $ij$ th position implies forecasting error between models  $i$  and  $j$  are significantly different at 5% alpha level.

Table 2

P-values of the Bonferroni Dunn's test for pairwise comparison of models' backcast error distributions.

Models (i, j)	ARIMA	KNN	NNAR	GRNN	MLP	ELM
KNN	0.789					
NNAR	1.000	0.000				
GRNN	1.000	0.411	0.247			
MLP	1.000	0.113	0.499	1.000		
ELM	0.000	0.722	0.000	0.000	0.000	
LSTM	1.000	1.000	0.373	1.000	1.000	0.010

\*  $p$ -value < 0.05 in  $ij$ th position implies backcasting error between models  $i$  and  $j$  are significantly different at 5% alpha level.

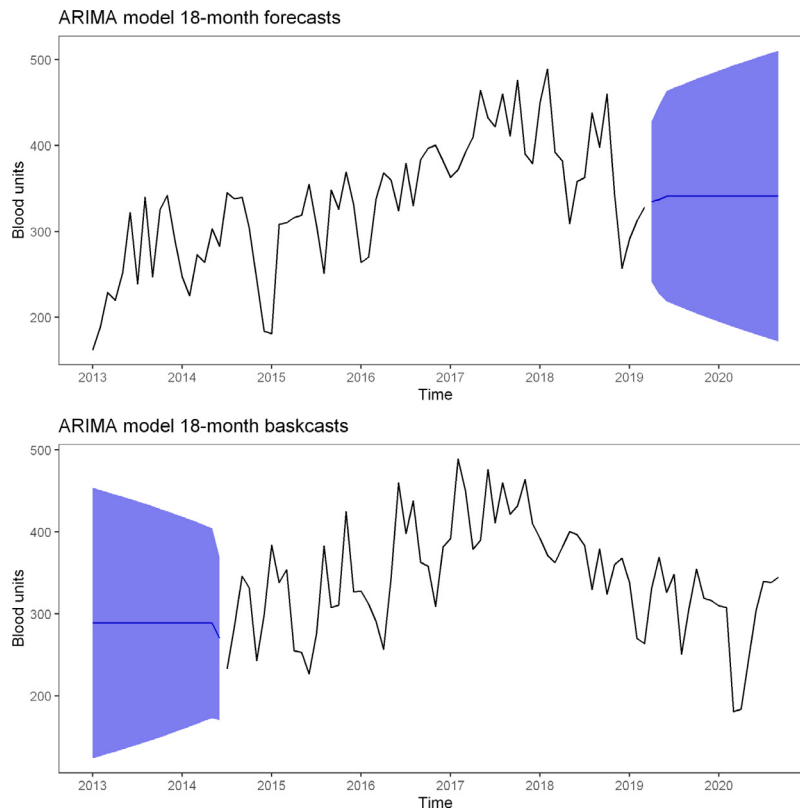


Fig. 6. A plot of the 18-month forecasts and backcasts from fitted ARIMA model.

**Table 3**

Pooled average prediction error (median MAPE) for empirical model comparison.

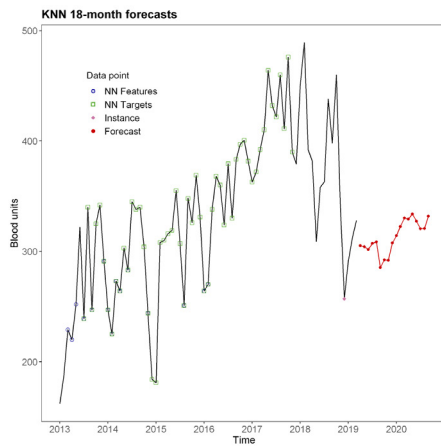
Time-series models	Median MAPE (%)	
	Forecast	Backcast
ARIMA	18.399	29.220
KNN	12.547	25.937
NNAR	17.634	33.959
GRNN	19.127	29.680
MLP	22.184	29.322
ELM	18.558	19.364
LSTM	18.942	28.212

\* The median MAPE quantifies the overall % prediction error.

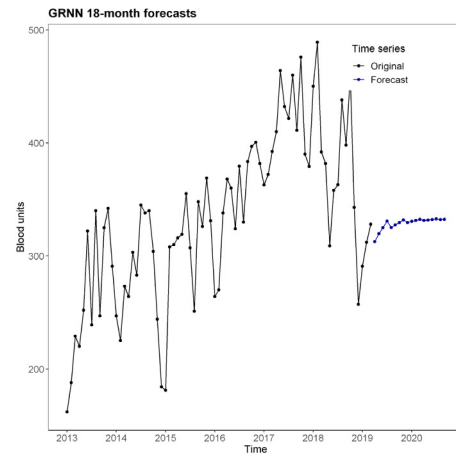
blood demand in a major hospital in Ghana, based on a short time-series data with unavailable past values; and contaminated with missing values and outliers. The classical non-seasonal ARIMA model was used as a baseline model for comparison with the ML models (KNN, NNAR, GRNN, MLP, ELM and LSTM algorithms), and a proposed rolling-origin strategy was employed for model evaluations. We also proposed a backcasting scheme to predict past values of lost data using time-series models provided the series is strictly stationary and time-reversible, such that its probabilistic structure remains invariant. The backcasts were obtained by reversing forecasts backwards

in time, and thus, the time-reversibility assumption was considered a necessary condition for the proposed backcasting scheme using time-series models. In addition, robust state-of-the-arts data preprocessing algorithms were adopted for data cleaning before the time-series modelling. Thus, the effect of different data preprocessing techniques and their respective performance on the predictive errors of the models were not significant concerns in this study.

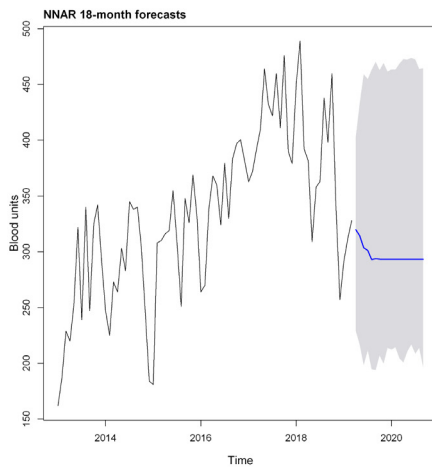
Previous studies have systematically shown that the traditional time-series methods such as ARIMA and exponential smoothing models usually outperform the complex machine learning algorithms (Makridakis et al., 2018). Nonetheless, Cerqueira, Torgo, and Soares (2019) discovered in their study that the prediction deficiency of ML models in such instances is predominantly due to potentially small training sample sizes. ML algorithms generally perform better with large training samples, though appropriate hyperparameter(s) tuning can also improve their predictive performance. Expressly, the two studies, as mentioned earlier, confirmed that traditional forecasting methods give better predictions relative to the ML models when up to only 144 observations are considered during model training (Cerqueira et al., 2019; Makridakis et al., 2018). However, the choice of the prediction-error measures, the type of out-of-sample evaluation tests employed, data preprocessing techniques used, and the



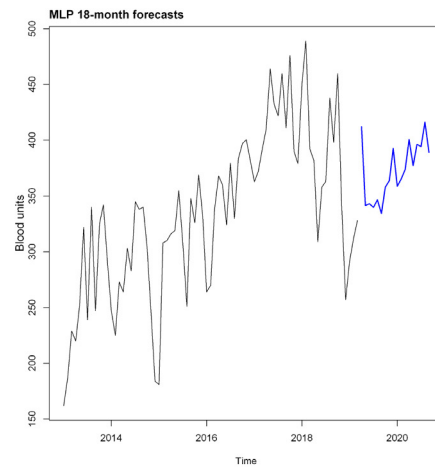
(a) KNN 18-month forecasts



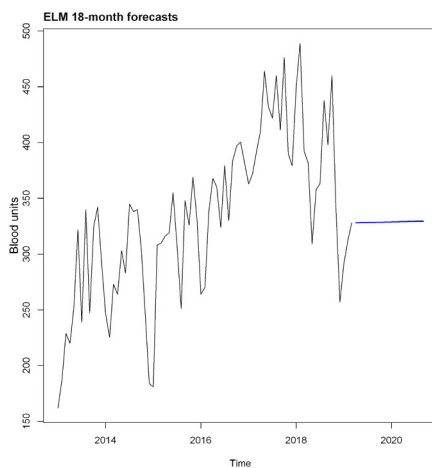
(b) GRNN 18-month forecasts



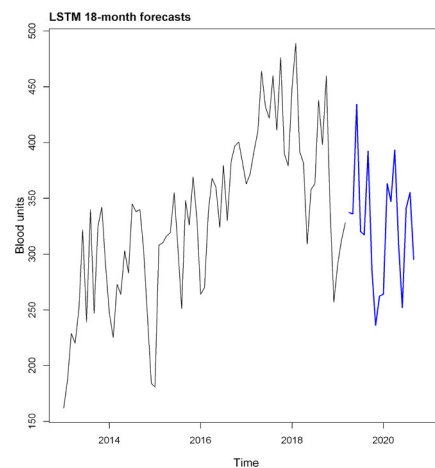
(c) NNAR 18-month forecasts



(d) MLP 18-month forecasts



(e) ELM 18-month forecasts

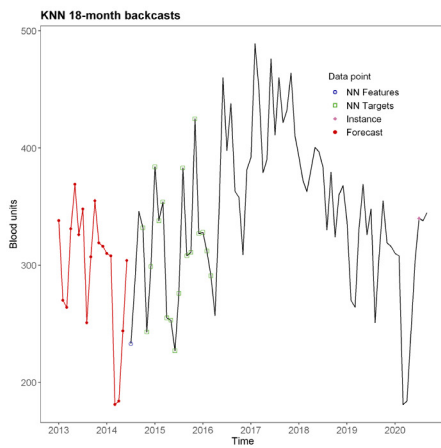


(f) LSTM 18-month forecasts

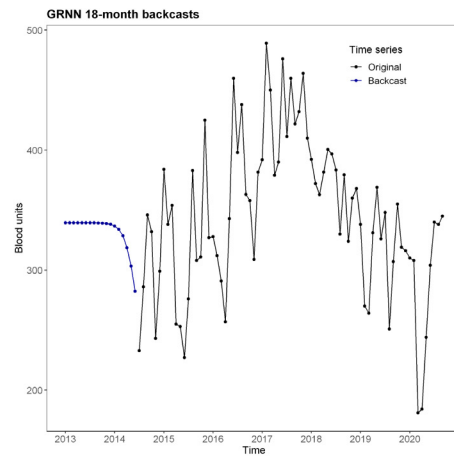
**Fig. 7.** Comparative plot of the 18-month blood demand forecasts from the fitted ML models.

direction of prediction (either forecasting or reverse forecasting) could also affect the validity of their conclusion, amongst other factors. In this study, the time-series

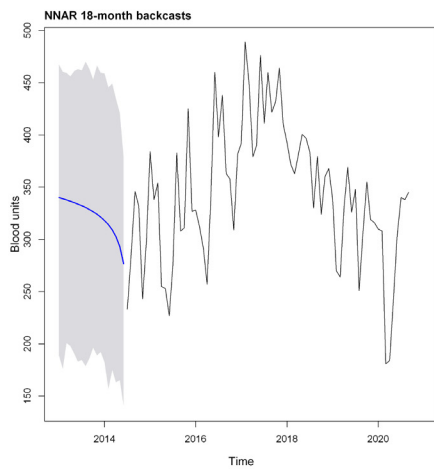
models were sequentially trained using different training samples of sizes ranging from 75 to 91 observations via a proposed rolling-origin strategy since the entire length of



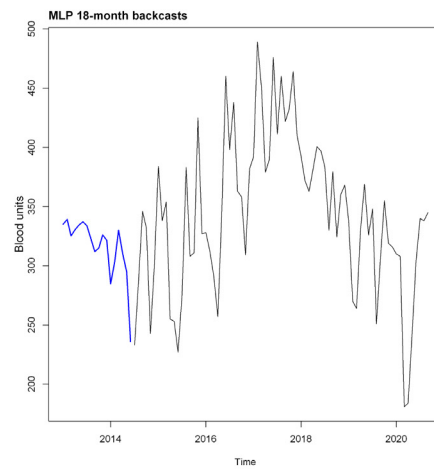
(a) KNN 18-month backcasts



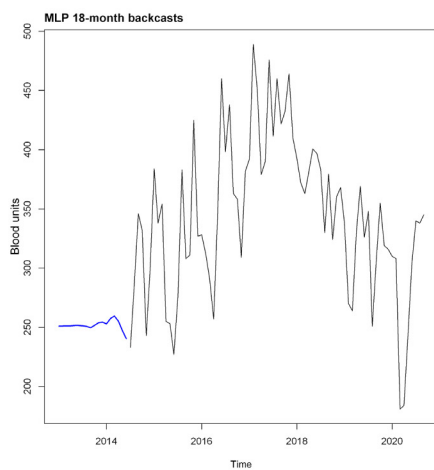
(b) GRNN 18-month backcasts



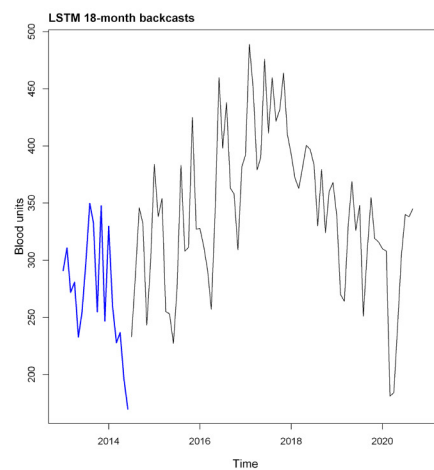
(c) NNAR 18-month backcasts



(d) MLP 18-month backcasts



(e) ELM 18-month backcasts



(f) LSTM 18-month backcasts

**Fig. 8.** Comparative plot of the 18-month blood demand backcasts from the fitted ML models.



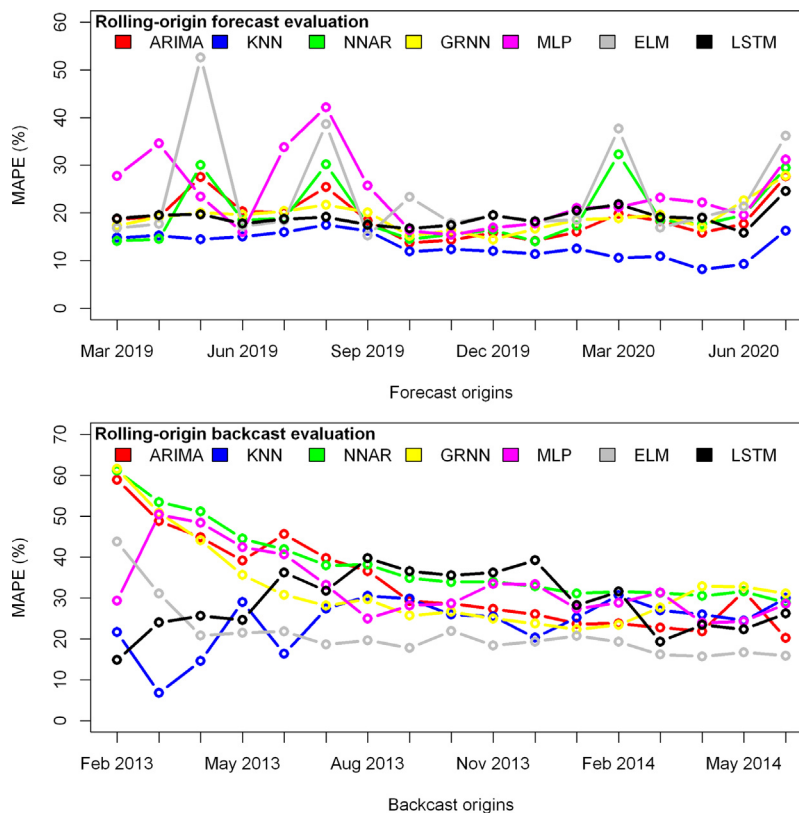


Fig. 9. Comparison between the prediction errors of the time-series models at different forecast (top) and backcast (bottom) origins.

the empirical data was 93. It was revealed that KNN and NNAR models outperformed the classical ARIMA model during blood demand forecasting, but there was no significant difference between their prediction errors across different forecasting origins, except for the KNN model. On the contrary, ELM models significantly had better backcasting power than the ARIMA model. Therefore, the direction of prediction (either forecasting or backcasting) can also affect the performance of machine learning algorithms and the traditional methods.

### 5.2. Managerial implications of the study

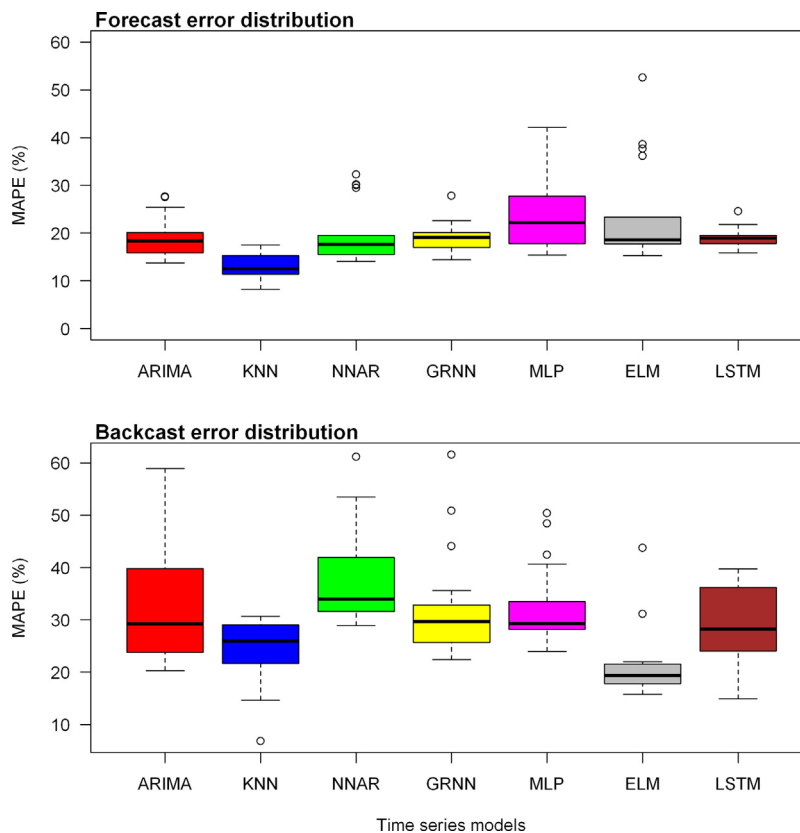
Blood centres are responsible for ensuring that blood products are available to promptly satisfy the demand from hospitals for transfusion to save lives. Reliable demand forecasting is thus the foundation of all blood supply chain planning and decision making (Silva Filho, Cezarino, & Salviano, 2012). However, there are often shortages in blood supply among blood banks or hospitals in Ghana due to low voluntary blood donation and associated challenges with the family-replenishment strategy (Stanger et al., 2012). The critical impeding factors for voluntary blood donations in Ghana are social and cultural beliefs, health risk concerns, and lack of proper education on sanitary blood donation (Harrington, 2013). Therefore, the rise in demand predominantly causes blood shortages, and thus, the need to identify time-series models that can efficiently predict blood demand over time

based on existing empirical data to avoid excess blood supply resulting in blood wastage. Unfortunately, small datasets like the one considered in this current study often exist where values of past years are either lost or completely unavailable.

Hence, we have demonstrated the application of machine learning algorithms and other time-series models like ARIMA to predict unavailable data of past years and make future predictions based on smaller sets of information or available data. Nonetheless, this study's main limitation is that the only available data on blood demand at the Tema General Hospital spanned from January 2013 to September 2020. Hence, it is inevitable that with more data, the conclusions would have been different concerning the predictive performance between the reference model (ARIMA) and the ML algorithms, but also regarding data preprocessing techniques and their sensitivity to the numbers of outliers.

## 6. Conclusion and recommendation

The study discovered that Extreme Learning Machine (ELM) and the K-Nearest Neighbour regression (KNN) algorithms are effective ML algorithms for predicting past values of unavailable blood demand data for blood centres and hospitals in Ghana via a reverse-forecast or backcast scheme. Time-series backcasting was only possible because the blood demand data was time-reversible and followed a Gaussian linear stochastic process. Even though



**Fig. 10.** Comparison of the prediction error distribution between ARIMA and the ML models.

K-Nearest Neighbour regression (KNN) and Neural Network Auto-regressive (NNAR) ML models outperformed the traditional ARIMA model in predicting future values based on the short time-series data, only KNN amongst the ML models, had a significantly lower forecasting error than the ARIMA model comparatively. Also, it was revealed that the direction of prediction (either forecasting or backcasting) could also affect the performance of the underlying time-series models.

The data correction method could significantly affect the predictive outcome of ML algorithms and other classical time-series models. Hence, we recommend that future studies investigate the effects of different data preprocessing techniques on the time-series models' predictive power for short or long series. Furthermore, the blood centres in Ghana should get proper database management systems to avoid data loss and outliers due to genuine recording errors. Future studies can also employ machine learning algorithms as a good alternative for backcasting past values of different time-series data with unavailable data of previous years.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Financial support

No funding was obtained for the study.

### Data availability

Data is made publicly available on GitHub and can be accessed together with all R scripts via the URL link: <https://github.com/twumasiclement/TimeSeries-Forecasting>.

### References

- Ahmar, A. S., Guritno, S., Abdurakhman, Rahman, A., Awi, Alimuddin, et al. (2018). Modeling data containing outliers using ARIMA additive outlier (ARIMA-AO). arXiv.
- Ahmed, N. K., Atiya, A. F., Gayar, N. E., & El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, 29(5–6), 594–621.
- Alajrami, E., Abu-Nasser, B. S., Khalil, A. J., Musleh, M. M., Barhoom, A. M., & Naser, S. A. (2019). Blood donation prediction using artificial neural network.
- Ali, A., Auvinen, M.-K., & Rautonen, J. (2010). Blood donors and blood collection: The aging population poses a global challenge for blood services. *Transfusion*, 50(3), 584–588.
- Allen, D. M. (1974). The relationship between variable selection and data augmentation and a method for prediction. *Technometrics*, 16(1), 125–127.
- Alloja, J., Espenberg, K., & Kiivet, R. (2012). The analysis of cost-effectiveness of the optimal system of the blood service. *Technical report*, (pp. 1–107). University of Tartu: Ordered by Ministry of Social Affairs.
- Alpaydin, E. (2020). *Introduction to machine learning*. MIT Press.

- Arnold, T. B. (2017). KerasR: R interface to the keras deep learning library. *Journal of Open Source Software*, 2(14), 296.
- Arvan, M., Tavakoli-Moghadam, R., & Abdollahi, M. (2015). Designing a bi-objective and multi-product supply chain network for the supply of blood. *Uncertain Supply Chain Management*, 3(1), 57–68. <http://dx.doi.org/10.5267/j.uscm.2014.8.004>.
- Asael Alonzo Matamoros, I., & Nieto-Reyes, A. (2020). An R package for normality in stationary processes. arXiv e-prints arXiv-2009.
- Baesler, F., Nemeth, M., Martínez, C., & Bastías, A. (2014). Analysis of inventory strategies for blood components in a regional blood center using process simulation. *Transfusion*, 54(2), 323–330. <http://dx.doi.org/10.1111/trf.12287>.
- Beliën, J., & Forcé, H. (2012). Supply chain management of blood products: A literature review. *European Journal of Operational Research*, 217(1), 1–16.
- Bibri, S. E. (2018). Backcasting in futures studies: a synthesized scholarly and planning approach to strategic smart sustainable city development. *European Journal of Futures Research*, 6(1), 1–27.
- Bontempi, G., Ben Taieb, S., & Le Borgne, Y. A. (2013). Machine learning strategies for time series forecasting. 138 LNBIP. In *Lecture notes in business information processing* (pp. 62–77). Springer Verlag. [http://dx.doi.org/10.1007/978-3-642-36318-4\\_3](http://dx.doi.org/10.1007/978-3-642-36318-4_3).
- Bontempi, G., Birattari, M., & Bersini, H. (1999). Local learning for iterated time series prediction. In *International conference of machine learning* (pp. 32–38). Citeseer.
- Bontempi, G., & Taieb, S. B. (2011). Conditionally dependent strategies for multiple-step-ahead prediction in local learning. *International Journal of Forecasting*, 27(3), 689–699.
- Bosnes, V., Aldrin, M., & Heier, H. E. (2005). Predicting blood donor arrival. *Transfusion*, 45(2), 162–170.
- BuHamra, S., Smaoui, N., & Gabr, M. (2003). The Box-Jenkins analysis and neural networks: prediction and time series modelling. *Applied Mathematical Modelling*, 27(10), 805–815.
- Caporin, M., & Sartore, D. (2011). Methodological aspects of time series back-calculation. *SSRN Electronic Journal*, (56), <http://dx.doi.org/10.2139/ssrn.950923>.
- Cerqueira, V., Torgo, L., & Soares, C. (2019). Machine learning vs statistical methods for time series forecasting: Size matters. arXiv preprint arXiv:1909.13316.
- Chaiwuttisak, P., Smith, H., Wu, Y., Potts, C., Sakuldamrongpanich, T., & Pathomsiri, S. (2016). Location of low-cost blood collection and distribution centres in thailand. *Operations Research for Health Care*, 9, 7–15. <http://dx.doi.org/10.1016/j.orhc.2016.02.001>.
- Chen, C., & Liu, L.-M. M. (1993). Forecasting time series with outliers. *Journal of Forecasting*, 12(1), 13–35. <http://dx.doi.org/10.1002/for.3980120103>.
- Constantino, A., Garcia, A., & Sawitzki, G. (2020). Package “NonlinearT-series”.
- Critchfield, G., Connelly, D., Ziehwein, M., Olesen, L., Nelson, C., & Scott, E. (1985). Automatic prediction of platelet utilization by time series analysis in a large tertiary care hospital. *American Journal of Clinical Pathology*, 85(5), 627–631.
- Custer, B. J., Sullivan, S. D., Hazlet, T. K., Ramsey, S. D., Hirschler, N. V., Murphy, E. L., et al. (2004). Quantifying losses to the donated blood supply due to donor deferral and miscollection. *Transfusion*, 44(10), 1417–1426.
- Da Veiga, C. P., Da Veiga, P. R., Catapan, A., Tortato, U., & Da Silva, W. V. (2014). Demand forecasting in food retail : a comparison between the Holt-Winters and ARIMA models. *WSEAS Transactions on Business Economics*, 11, 608–614.
- Darwiche, M., Feuilloley, M., Bousaleh, G., & Schang, D. (2010). Prediction of blood transfusion donation. In *2010 fourth international conference on research challenges in information science (RCIS)* (pp. 51–56). IEEE.
- De Kort, W., Janssen, M., Kortbeek, N., Jansen, N., van der Wal, J., & van Dijk, N. (2011). Platelet pool inventory management: Theory meets practice. *Transfusion*, 51(11), 2295–2303. <http://dx.doi.org/10.1111/j.1537-2995.2011.03190.x>.
- Dillon, M., Oliveira, F., & Abbasi, B. (2017). A two-stage stochastic programming model for inventory management in the blood supply chain. *International Journal of Production Economics*, 187, 27–41.
- Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods*. Oxford University Press.
- Faraway, J., & Chatfield, C. (1998). Time series forecasting with neural networks: a comparative study using the air line data. *Journal of the Royal Statistical Society. Series C. Applied Statistics*, 47(2), 231–250.
- Fortsch, S. M., & Khapalova, E. A. (2016). Reducing uncertainty in demand for blood. *Operations Research for Health Care*, 9, 16–28.
- Francisco Martinez, M. (2019). Package “tsfgrnn” title time series forecasting using GRNN. [http://dx.doi.org/10.1007/978-3-030-20521-8\\_17](http://dx.doi.org/10.1007/978-3-030-20521-8_17).
- Friedman, J., Hastie, T., Tibshirani, R., et al. (2001). Vol. 1, *The elements of statistical learning*. (10), Springer series in statistics New York.
- Ghana Health Service (2014). Blood bank outlook | tema general hospital | ghana health service.
- Gopika, P., Sowmya, V., Gopalakrishnan, E., & Soman, K. (2020). Transferable approach for cardiac disease classification using deep learning. In *Deep learning techniques for biomedical and health informatics* (pp. 285–303). Elsevier.
- Harrington, A. H. (2013). Blood banks in kumasi, ghana: Social barriers preventing volunteer blood donations.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.
- Hinich, M. J. (2005). *Time series analysis by state space methods*. Taylor & Francis.
- Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2006). Extreme learning machine: theory and applications. *Neurocomputing*, 70(1–3), 489–501.
- Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.
- Hyndman, R. J., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., et al. (2020). Package “forecast”. <https://cran.r-project.org/web/packages/forecast/forecast.pdf>.
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688.
- Jayaweera, C., & Aziz, N. (2018). Development and comparison of extreme learning machine and multi-layer perceptron neural network models for predicting optimum coagulant dosage for water treatment. *Journal of Physics: Conference Series*, 1123(1), Article 012032.
- Jittamai, P., & Boonyanusith, W. (2014). Risk assessment in managing the blood supply chain. In *Next generation supply chains: trends and opportunities. proceedings of the hamburg international conference of logistics (HICL)*. Vol. 18 (pp. 447–468). Berlin: epubli GmbH.
- Kantz, H., & Schreiber, T. (2004). Vol. 7, *Nonlinear time series analysis*. Cambridge University Press.
- Katsaliaki, K., & Brailsford, S. C. (2007). Using simulation to improve the blood supply chain. *Journal of the Operational Research Society*, 58(2), 219–227.
- Khalidi, R., El, A., Chihe, R., Khald, R., El, A., & Chihe, R. (2019). Forecasting of weekly patient visits to emergency department : real forecasting of weekly patient visits to emergency department : real case study case study. *Procedia Computer Science*, 148, 532–541. <http://dx.doi.org/10.1016/j.procs.2019.01.026>.
- Khalidi, R., El Afia, A., Chiheb, R., & Faizi, R. (2017). Artificial neural network based approach for blood demand forecasting: Fez transfusion blood center case study. In *Proceedings of the 2nd international conference on big data, cloud and applications* (pp. 1–6).
- Khashei, M., Bijari, M., & Hejazi, S. R. (2012). Combining seasonal ARIMA models with computational intelligence techniques for time series forecasting. *Soft Computing*, 16(6), 1091–1105. <http://dx.doi.org/10.1007/s00500-012-0805-9>.
- Kopach, R., Balcioglu, B., & Carter, M. (2008). Tutorial on constructing a red blood cell inventory management system with two demand rates. *European Journal of Operational Research*, 185(3), 1051–1059.
- Kourentzes, N. (2017). Nnfor: Time series forecasting with neural networks. R Package.
- López-de Lacalle, J. (2016). Tsoutliers R package for detection of outliers in time series. CRAN, R Package.
- Leung, M. T., Chen, A.-S., & Daouk, H. (2000). Forecasting exchange rates using general regression neural networks. *Computers & Operations Research*, 27(11–12), 1093–1110.
- Li, N., Chiang, F., Down, D. G., & Heddle, N. M. (2021). A decision integration strategy for short-term demand forecasting and ordering for red blood cell components. *Operations Research for Health Care*, 29, Article 100290.

- Maia, L. S. A., De Carvalho, F. A. T., & Ludermit, T. B. (2008). Neurocomputing forecasting models for interval-valued time series. *Neurocomputing*, 71, 3344–3352. <http://dx.doi.org/10.1016/j.neucom.2008.02.022>.
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and machine learning forecasting methods: Concerns and ways forward. *PLoS One*, 13(3), Article e0194889.
- Martínez, F., Frías, M. P., Charte, F., & Rivera, A. J. (2019a). Time series forecasting with KNN in R: the tsfkn package. *The R Journal*.
- Martínez, F., Frías, M. P., Pérez, M. D., & Rivera, A. J. (2019b). A methodology for applying k-nearest neighbor to time series forecasting. *Artificial Intelligence Review*, 52(3).
- Moritz, S., & Bartz-Beielstein, T. (2017). Imputets: Time series missing value imputation in R. *R Journal*, 9(1), 207–218. <http://dx.doi.org/10.32614/rj-2017-009>.
- Nene, D., Olayemi, E., & Asamoah-Akuoko, L. (2015). Application of information technology by blood banking and transfusion facilities in Ghana. *Africa Sanguine*, 17(2), 7–10.
- Ollech, D., & Webel, K. (2020). A random forest-based approach to identifying the most informative seasonality tests: Number 55/2020, Bundesbank Discussion Paper.
- Osorio, A. F., Brailsford, S. C., & Smith, H. K. (2015). A structured review of quantitative models in the blood supply chain: a taxonomic framework for decision-making. *International Journal of Productions Research*, 53(24), 7191–7212.
- Pankratz, A. (2009). vol. 224, *Forecasting with univariate Box-Jenkins models: concepts and cases*. John Wiley & Sons.
- Papacharalampous, G., Tyralis, H., & Koutsoyiannis, D. (2018). Univariate time series forecasting of temperature and precipitation with a focus on machine learning algorithms: a multiple-case study from Greece. *Water Resources Management*, 32(15), 5207–5239. <http://dx.doi.org/10.1007/s11269-018-2155-6>.
- Pereira, A. (2004). Performance of time-series methods in forecasting the demand for red blood cell transfusion. *Transfusion*, 44(5), 739–746.
- Phdungsilp, A. (2011). Futures studies' backcasting method used for strategic sustainable city planning. *Futures*, 43(7), 707–714.
- Pierskalla, W. P. (2005). Supply chain management of blood banks. In *Operations research and health care* (pp. 103–145). Springer.
- R Core Team (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, <http://www.R-project.org/>.
- Raeth, C., & Monetti, R. (2009). Surrogates with random Fourier phases. In *Topics on chaotic systems: selected papers from chaos 2008 international conference* (pp. 274–285). World Scientific.
- Rajendran, S., & Ravindran, A. R. (2019). Inventory management of platelets along blood supply chain to minimize wastage and shortage. *Computers & Industrial Engineering*, 130, 714–730.
- Rajendran, S., & Srinivas, S. (2020). Hybrid ordering policies for platelet inventory management under demand uncertainty. *IIE Transactions on Healthcare Systems Engineering*, 10(2), 113–126.
- Ramaekers, K., & Janssens, G. K. (2014). Optimal policies for demand forecasting and inventory management of goods with intermittent demand. *Journal of Applied Operational Research*, 6(2), 111–123.
- Rasyid, A., & Adhiutama, A. (2014). Forecasting the demand for PT Telkom's internet service using time series analysis. In *Proceedings of 7th Asia-Pacific business research conference 25 - 26 august 2014, Bayview hotel, Singapore* (pp. 1–18).
- Robinson, J. B. (1990). Futures under glass: a recipe for people who hate to predict. *Futures*, 22(8), 820–842.
- Rosenblatt, F. (1961). Principles of neurodynamics. perceptrons and the theory of brain mechanisms. *Technical report*, Cornell Aeronautical Lab Inc Buffalo NY.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1985). Learning internal representations by error propagation. *Technical report*, California Univ San Diego La Jolla Inst for Cognitive Science.
- Ryttilä, J. S., & Spens, K. M. (2006). Using simulation to increase efficiency in blood supply chains. *Management Research News*.
- Schreiber, T., & Schmitz, A. (2000). Surrogate time series. *Physica D: Nonlinear Phenomena*, 142(3–4), 346–382.
- Sharifdoust, M., & Mahmoodi, S. (2013). On time reversibility of linear time series. *Journal of Mathematical Extension*, 6, 33–47.
- Shih, H., & Rajendran, S. (2019). Comparison of time series methods and machine learning algorithms for forecasting Taiwan blood services foundation's blood supply. *Journal of Healthcare Engineering*, 2019.
- Shokouhifar, M., Sabbaghi, M. M., & Pilevari, N. (2021). Inventory management in blood supply chain considering fuzzy supply/demand uncertainties and lateral transshipment. *Transfusion and Apheresis Science*, Article 103103.
- Silva-Filho, O. S., Carvalho, M. A., Cezarino, W., Salva, R., & Silviano, G. (2013). Demand forecasting for blood components distribution of a blood supply chain. In *6th IFAC conference on management and control of production and logistics* (pp. 565–571). The International Federation of Automatic Control, <http://dx.doi.org/10.3182/20130911-3-BR-3021.00092>.
- Silva Filho, O. S., Cezarino, W., & Salviano, G. R. (2012). A decision-making tool for demand forecasting of blood components. *IFAC Proceedings Volumes*, 45(6), 1499–1504.
- Simon, S. (2007). A comparison of the forecasting ability of ARIMA models. *Journal of Property Investment & Finance*, 25(3), 223–240. <http://dx.doi.org/10.1108/14635780710746902>.
- Stanger, S. H., Wilding, R., Yates, N., & Cotton, S. (2012). What drives perishable inventory management performance? Lessons learnt from the UK blood supply chain. *Supply Chain Management: An International Journal*.
- Taieb, S. B., Bontempi, G., Atiya, A. F., & Sorjamaa, A. (2012). A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. *Expert Systems with Applications*, 39(8), 7067–7083.
- Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting*, 16(4), 437–450.
- Tashman, L. J., & Kruk, J. M. (1996). The use of protocols to select exponential smoothing procedures: A reconsideration of forecasting competitions. *International Journal of Forecasting*, 12(2), 235–253.
- Teräsvirta, T., Lin, C.-F., & Granger, C. W. (1993). Power of the neural network linearity test. *Journal of Time Series Analysis*, 14(2), 209–220.
- Teviu, E., Aikins, M., Abdulai, T., Sackey, S., Boni, P., Afari, E., et al. (2012). Improving medical records filing in a municipal hospital in Ghana. *Ghana Medical Journal*, 46(3), 136.
- The Lancet (2005). Blood supply and demand. *The Lancet*, 365(9478), 2151. [http://dx.doi.org/10.1016/S0140-6736\(05\)66749-9](http://dx.doi.org/10.1016/S0140-6736(05)66749-9).
- Thoplan, R. (2014). Simple v/s sophisticated methods of forecasting for mauritius monthly tourist arrival data. *International Journal of Statistics and Applications*, 4(5), 217–223.
- Vabalas, A., Gowen, E., Poliakov, E., & Casson, A. J. (2019). Machine learning algorithm validation with a limited sample size. *PLoS One*, 14(11).
- Van Dijk, N., Haijema, R., Van Der Wal, J., & Sibinga, C. S. (2009). Blood platelet production: a novel approach for practical optimization. *Transfusion*, 49(3), 411–420.
- Wakefield, K. (2013). A guide to machine learning algorithms and their application. Undated, SAS. Com, [https://www.sas.com/en\\_gb/insights/articles/analytics/machine-learning-algorithms.html](https://www.sas.com/en_gb/insights/articles/analytics/machine-learning-algorithms.html).