Artificial Neural Network Based Approach for Blood Demand Forecasting: Fez Transfusion Blood Center Case Study

Rohaifa Khaldi
ENSIAS
Mohammed V University, Rabat, Morocco
rohaifa.khaldi@gmail.com

Abdellatif El Afia
ENSIAS
Mohammed V University, Rabat, Morocco
a.elafia@um5s.net.ma

Raddouane Chiheb
ENSIAS
Mohammed V University, Rabat, Morocco
r.chiheb@um5s.net.ma

Rdouan Faizi
ENSIAS
Mohammed V University, Rabat, Morocco
r.faizi@um5s.net.ma

ABSTRACT
Blood demand and supply management are considered one of the major components of a healthcare supply chain, since blood is a vital element in preserving patient's life. However, forecasting it faces several challenges including frequent shortages, and possible expiration caused by demand uncertainty of hospitals. This uncertainty is mainly due to high variability in the number of emergency cases. Thereupon, this investigation presents a real case study of forecasting monthly demand of three blood components, using Artificial Neural Networks (ANNs). The demand of the three blood components (red blood cells (RBC), plasma (CP) and platelets (PFC)) and other observations are obtained from a central transfusion blood center and a University Hospital. Experiments are carried out using three networks to forecast each blood component separately. Last, the presented model is compared with ARIMA to evaluate its performance in prediction. The results of this study depict that ANN models overcomes ARIMA models in demand forecasting. Thus high ANN models can be considered as a promising approach in forecasting monthly blood demand.

CCS Concepts
• **Applied computing → Operations research → Forecasting** • **Computing methodologies → Machine learning → Machine learning approaches → Neural networks.**

Keywords
Artificial neural networks; blood components demand; forecasting; blood supply chain.

1. INTRODUCTION
Over the past several years, developments in the field of healthcare have been focusing on enhancing supply chains [1, 2]. As explained in [5], the ultimate goal is to decrease wastage and shortage of resources, and to reduce healthcare overall costs while improving the public health, patients' safety, and service level quality.

Copyright 2017 ACM ISBN 978-1-4503-4852-2/17/03…$15.00.
DOI: http://dx.doi.org/10.1145/3090354.3090415

One of the most important and critical components of healthcare supply chain is blood demand management.
Since it feeds hospitals with blood components, which are considered as the vital elements of human life. In order to meet the need of their regular and irregular patients, blood components are provided periodically to hospitals by blood centers.

Blood delivery from blood center to hospitals is done according to different shift of time (daily, weekly, monthly,…), most of regular shipments in the urban areas are done each day. However, in the rural areas they may be done each week. Blood center and hospitals should maintain an optimal inventory levels in their blood banks [9].

Services provided by a blood center are affected by many disruptions that impede the blood channel. One type of these disruptions occurs during the blood collection process [1]. Notably while looking for uncontaminated blood or while looking for donors with a rare type of blood. Another type of these disruptions may be caused by the wastage and shortage of the different blood components [1, 9]. In fact, excess of blood components may lead to blood perishability, and hence loss of financial resources. On the other hand, blood shortages may lead to complicated medical situations or to even loss of human life. We point out that these disruptions are mainly due the stochastic aspect of blood demand, which is caused by the high variability of hospital demand.

Indeed, demand of blood is a major source of uncertainty, that's why it is fundamental to understand past, present and future demand of a wide variety of blood product. Accordingly, managers should focus on the implementation of efficient techniques. This consist of forecasting in order to control shortage and availability of blood components, reduce wastage and product expiration, decrease blood inventory levels, and control other related costs [9]. Hence the importance of using accurate forecasting systems [1, 3].

Correspondingly, this study contributes to the existing literature by presenting three ANN optimal models to forecast monthly demand of three blood components. The usage of this computational intelligence paradigm aims to increase the accuracy of forecasting. According to [7] the certainty and the accuracy of demand forecasting is the foundation of each healthcare supply chain planning. For this reason, the results of ANN models are evaluated using statistical indicators and benchmarked using the results of ARIMA models.

For our best knowledge, this investigation fills the gap in the literature by constructing a model based on artificial neural network to forecast blood demand components. This model will help blood centers and hospitals to identify the amount of blood bags to be delivered to hospitals in the future. As well as to diminish the uncertainty of blood demand, and to reduce costs and blood wastage.

The remainder of this paper is structured as follows: Section 2 spotlights the background literature while Section 3 presents the methodology used for forecasting. Section 4 applies the proposed model in blood supply chain, while Section 5 evaluates the model and discusses the results. Finally, Section 6 concludes the study and provides ideas for future research.

2. LITERATURE REVIEW

Different researchers have provided many solutions to bring down blood wastage. These solutions deal with managing inventory, delivery, operations and planning. [18] explore policies of blood management in networks so that to reduce costs and bring up the efficiency for the Canadian Blood Services. [19] present perceivable literature review pertaining donor arrivals, waste reduction, and ordering policies. [20] use a logistic regression model to forecast the arrivals of blood donors.

Indeed, there exist a number of investigations dealing with forecasting inside blood supply chain [10, 19, 20]. However, applications of forecasting dealing with blood demand are a rarity. [10] propose exponential smoothing techniques to forecast transfusions at hospitals. [4] introduce multiple practical approaches to predict blood demand, and teste several forecasting models such as the Naïve, Moving Average (MA), Exponential Smoothing (ES), and Time Series Decomposition (TSD) models. Then they compared them to ARIMA model. The results showed that ARIMA model is the optimal choice to forecast demand, because it provides accurate forecasts for blood demand.

These econometric techniques are limited because they are based solely on the previous values of endogenous variable without any other exogenous variables. They don't take into consideration other factors that may impact the demand of blood.

The present study approaches the problem of blood demand forecasting using ANN model. The major features of neural networks are their ability to perform a complex nonlinear mapping. This mapping goes from multidimensional input space onto multidimensional output space, without knowing any relationships between input and output spaces [14]. Another feature, is its ability to deal with real world phenomena by withstanding data with errors, and using sequential training procedures to adapt themselves to new data. Even though, the mathematics incorporated in neural networking is not simple, a user can easily acquire at least an operational understanding of neural network structure and function [11].

3. THE PROPOSED APPROACH

Artificial neural networks (ANN) is an intelligent system inspired by the human nervous system. ANNs are very good with fitting problems, with enough neurons ANNs can fit any data with arbitrary accuracy.

Neural Network links a set of input nodes $x_i = (x_1, x_2, \ldots, x_n)$ existing in the input layer with a set of one or more output nodes $y_j = (y_1, y_2, \ldots, y_m)$ existing in the output layer through an intermediate hidden layer.

Nodes in each layer are activated once they reach the layer threshold value θ_i. This matching is realized by finding an unknown function h.

$$\mathrm{y_j} = \mathrm{h}(\mathrm{x_1}, \mathrm{x_2}, \ldots, \mathrm{x_n}) \tag{1}$$

$i \in [1, n]$ and $j \in [1, m]$

To implement artificial neural networks the user should follow the plotted flow chart in Figure 1. First, data should be collected and partitioned into input dataset and desired output dataset. Second, the user should build and design his network by choosing the type of learning: supervised learning, unsupervised learning or reinforcement learning. As well as by fixing the network parameters, for example net input, transfer function, learning function, learning rate, number of neurons in each layer, etc.

Then, the dataset should be preprocessed using either normalization or standardization. After, it should be divided on training data, validation data and testing data. In fact, the training dataset is used to identify the values of weights and biases of the network. While the validation dataset is employed to analyze weights and biases, so as to measure the capacity of network generalization, and to interrupt training when overfitting occurs [15]. Whereas, the testing dataset is used to validate weights and biases participating in the stopping criterion, and to evaluate the network performance on new datasets [6].

Finally, the network should be simulated, and if necessary its settings should be modified until obtaining good results.

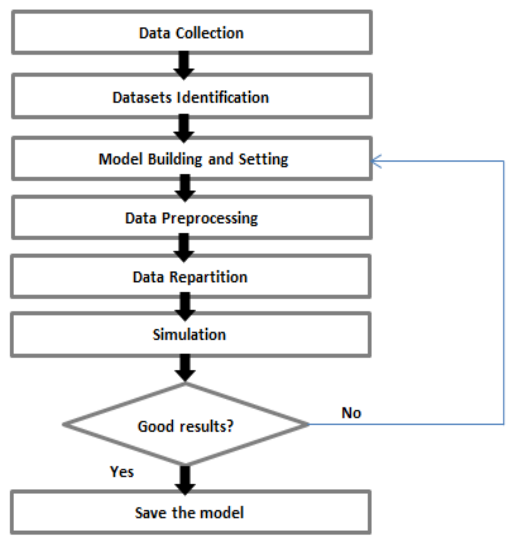

Figure 1. The flow-chart of ANNs

4. EXPERIMENT

4.1 Study Design

This research presents a retrospective study using aggregated data from the year 2011 to 2015, conducted in Morocco in the region of Fez-Meknes.

This study was conducted using data (such as number of accidents, surgeries, and admissions) extracted from the records of services of two hospitals working under the University Hospital Hassan II of Fez (CHU Hassan II) including: Specialty Hospital and Mother-Child Hospital. Data of blood demand was also collected from the regional center of blood transfusion (CRTS) of Fez city.

In fact the University Hospital Hassan II of Fez was chosen because it is the bigger user of blood components (RBC, CP and PFC) in Fez city.

The University Hospital includes five hospitals: Specialty Hospital, Mother-Child Hospital, Oncology Hospital, Otorhinolaryngology Hospital, and Psychiatric Hospital. From these five sub-hospitals only three (Specialty Hospital, Mother-Child Hospital and Oncology Hospital) demand the three blood components, but in our case we will not include the demand of the Oncology Hospital.

4.2 Data Collection

Data for this analysis were collected from different data warehouses of the CHU services and the regional blood center. The daily data of blood demand was collected from the year of 2010 to 2015, then aggregated into monthly demand because the available information in the CHU is monthly. Table 1 shows the distribution of blood demand components over the years.

Table 1. The distribution of blood demand components

Blood type	**Years**					
	2010	**2011**	**2012**	**2013**	**2014**	**2015**
RBC	11560	10892	12595	12472	13835	12335
CP	6716	6363	6068	6422	9379	8030
PFC	5220	6291	6768	5241	8030	5379

4.3 Datasets Identification

To forecast the demand of blood components, we have looked for the factors that may impact the blood demand variation in the CHU such as surgeries, emergency cases, accidents, when the patient is anemic, in deliveries, in hemorrhage cases and during the process of dialysis. Ten factors were selected as inputs (independent variables) in this investigation as shown in Table 2.

Table 2. Blood input parameters

Input 1	Admissions
Input 2	Days of hospitalization
Input 3	Scheduled surgeries
Input 4	Non-scheduled surgeries
Input 5	Emergency passages
Input 6	Public roadway accidents
Input 7	Dialysis
Input 8	Deliveries at risk
Input 9	Caesareans
Input 10	Hemorrhages

The first input represents the admissions, which is the number of patients admitted to the hospital. The second input represents the days of hospitalization, which is the number of stays of patients inside the hospital. The third input presents the scheduled surgeries, which represent the number of surgeries that have been already programmed within the hospital. The fourth input depicts the non-scheduled surgeries, which represent the number of emergency surgeries in the hospital. The fifth input presents the emergency passages, representing the number of emergency cases in the hospital. The sixth input presents the public roadway accidents, which represent the number of accidents occurred in the public roadway. The seventh input depicts the dialysis, which present the number of dialysis occurred in the hospital. The eighth input presents the deliveries at risk, representing the number of childbirths that had been occurred in tough conditions within the hospital. The ninth input is the caesareans, which is the number of caesareans occurred within the hospital. The tenth input represents the number of hemorrhages cases in the hospital.

Table 3 presents the output parameters used in each network of this investigation, including the number of red blood cells, plasma and platelets demanded by the CHU.

Table 3. Blood output parameters

Output of Network 1	Demand of red blood cells
Output of Network 2	Demand of plasma
Output of Network 3	Demand of platelets

4.4 Model Building and Setting

This investigation is based on three models having the same input variables and different output variables (demand of red blood cells, plasma and platelets).

The neural networks proposed for this research are multilayer feedforward neural networks operating under supervised learning; they consist of three layers including one input layer, one hidden layer and one output layer. Indeed more hidden neurons provide better learning results. However, some previous studies assert that increasing the number of hidden neurons couldn't enhance the learning results, but could only increase the process of learning time [12]. In this study, hidden neurons were fixed through trial and error, therefore once the number of hidden neurons exceeds 15 the improvements become marginal to null. Hence, the structure of the three models is 10-15-1. Also, ANN parameters were chosen based on trial and errors. In this study, the learning rate is 0.01 and the momentum is 0.001.

The input layer, the hidden layer, and the output layer are fully interconnected. The weights of the connections and the biases are initialized randomly by the system, and adjusted through the learning process.
In order to map the sum of weighted inputs in the range of $[-1,1]$, tangent sigmoid transfer function is employed in the hidden layer of the three neural networks as given by Eq.2.

$$\mathrm{f(t)} = \frac{2}{1+\mathrm{e}^{-2\mathrm{t}}} - 1 \qquad (2)$$

This function is often used in neural network applications thanks to its insensitivity to noise, its function features and its continuous derivatives [13]. Otherwise, linear transfer function is used for output layer.

Modeling real world problems is based on adjusting a set of node parameters called weights, to form mappings from a set of given input variables to a set of desired output variables. These weights adjustments are made using the Levenberg–Marquardt learning algorithm that seeks to minimize the quadratic error. This learning algorithm was chosen as learning algorithm for its fruitful application in ANN, for function approximation and prediction. Before training, weights have no meaning because they are set randomly, however, after training they hold meaningful information [16].

4.5 Data Preprocessing

Before training the networks, inputs and targets were standardized by normalizing the mean and the standard deviation of the data set, in order to have zero mean and unity standard deviation. The following equation shows the standardization process:

$$x_n = \frac{x_i - \bar{x}}{\sigma} \quad (3)$$

Where $\mathbf{x_i}$ is the ith data point, $\boldsymbol{\sigma}$ and $\bar{\boldsymbol{x}}$ are the standard deviation and the mean respectively.

4.6 Data Repartition

Due to the insufficient data sets available for this study (72 observations) the training, validation and testing distributions was divided as follow (Table 4):

Table 4. Data Repartition

	Network 1	Network 2	Network 3
Training	80%	90%	90%
Testing	10%	5%	5%
Validation	10%	5%	5%

4.7 Performance Evaluation

To benchmark the models of ANN, prediction using autoregressive integrated moving average (ARIMA) was performed. Moreover, to evaluate the performance and the accuracy of the three networks, and to assess their quality of forecasting compared to ARIMA, their performance was analyzed by means of calculations of root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and the correlation coefficient (R), which are, respectively, shown in Eqs. (4) – (7):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(T_i - O_i)^2}{n}} \quad (4)$$

$$MAE = \frac{\sum_{i=1}^{n}|T_i - O_i|}{n} \quad (5)$$

$$MAPE(\%) = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{|T_i - O_i|}{T_i}\right) \times 100 \quad (6)$$

$$R = \sqrt{1 - \frac{\sum_{i=1}^{n}(T_i - O_i)^2}{\sum_{i=1}^{n} {T_i}^2}} \quad (7)$$

Where O_i are predicted values and T_i are real values.

To assess the performance of a predictive model, [17] recommended the following values of (R) statistical criterion:

- If a model provides |R| > 0.8, a strong correlation exists between the real values and the predicted values.
- If a model provides 0.2 < |R| < 0.8, a correlation exists between the real values and the predicted values.
- If a model provides |R| < 0.2, a weak correlation exists between the real values and the predicted values.

MAPE was used because we don't have any null demand and there is no negative value.

5. RESULTS AND DISCUSSION

Table 5 presents the numerical values obtained from the statistic indicators, used to evaluate the significance of the results provided by the different ANN and ARIMA models.

ARIMA models show that there is no trend and no seasonality effect in the first model which mean that the model is stationary. However, it indicates that second model exhibits seasonality behavior. While the last one display increasing trend with seasonality.

The results obtained imply that the evaluation process of performance is clearly dominated by ANN models. ANN models, particularly ANN2 exhibits a higher correlation coefficient (R), which mean that there exist a strong correlation between the real values and the predicted values. Indeed, RMSE values are higher for all the models. However, ANN models seem to have better performance and higher accuracy than ARIMA models. In fact, ANN2 delivers the lower RMSE, moreover, all ANN models indicate lower error measures than ARIMA models. For instance, MAE and MAPE values show that all forecasts are close to eventual outcomes for ANN models than for ARIMA models, with ANN3 is less accurate than other ANN models.

Figures 2, 3 and 4 depict the evolution of real and estimated values of the three types of blood demand (RBC, CP and PFC). They point out that all the ANN models express a slight under-forecasting of blood demand. This under-forecasting is a bit severe in ANN3, while it is soft in ANN1 and ANN2 where there is small difference between forecasted values and real ones. The intensity of this manifestation was provided by the values of MAPE: 2.06% for ANN1, 3.06% for ANN2 and 39.22% for ANN3.

The results of this analysis reveal that ANN is a better forecasting mechanism than ARIMA, and it can be considered as a competing alternative to predict future demands of blood components, because of its best performance and accuracy. However, ARIMA model enables the explanation of data and the interpretation of results, which are impossible when using ANN model. Because it is considered as "black box". Otherwise, the high values of RMSE appear due to data insufficiency, and due to the choice of data repartition, network training algorithm and network tuning parameters.

6. CONCLUSION AND FUTUTRE WORKS

The primary purpose of this paper is to explore the capabilities of ANN in forecasting blood demand. For this task, the proposed three models showed great accuracy in forecasting monthly demand of blood components than ARIMA models. They also showed a strong ability to deal with stochastic data characterized by non-linearity. For the best of our knowledge, this investigation seems to be the first initiative in incorporating artificial neural networks (ANNs) model in the forecasting system of the transfusion blood center. Subsequently, the results of this study can be used to help the staff of the transfusion blood center to forecast future demands of red blood cells, platelets and plasma. This is in order to avoid blood stock-outs and to prevent blood outdating. Otherwise, this investigation encouraged hospitals and blood centers to work cooperatively, to exchange and share proper information fundamental in blood forecasting. So that to enhance their quality service, to meet the needs of patients and to preserve their precious life.

Further studies can be accomplished to improve the efficiency of these models, by using feature selection in order to reduce the dimensionality of input set and to eliminate redundancy. As well as we recommend to increase the size of data set by using daily demand. Further, to use hybridization of machine learning algorithms, so as to improve the performance of ANNs.

Table 5. Statistical indicators of performance

	Models	RMSE	MAE	MAPE (%)	R
ARIMA1	(1,0,0)	242.37	179.63	9.71	0.36
ANN1	10-15-1	84.14	38.91	2.06	0.95
ARIMA2	$(1,0,0)(0,0,1)_{12}$	99.11	70.52	16.21	0.68
ANN2	10-15-1	39.37	13.41	3.06	0.96
ARIMA3	$(1,0,0)(0,1,0)_{12}$	151.88	93.07	42.95	0.18
ANN3	10-15-1	67.52	44.74	39.22	0.93

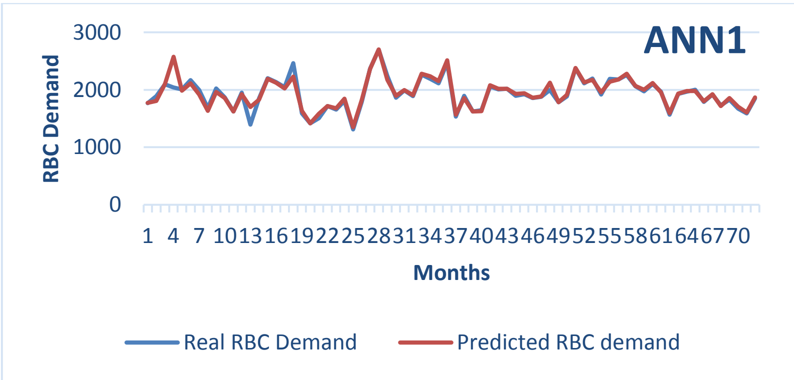

Figure 2. Predicted and Real RBC demand of ANN1

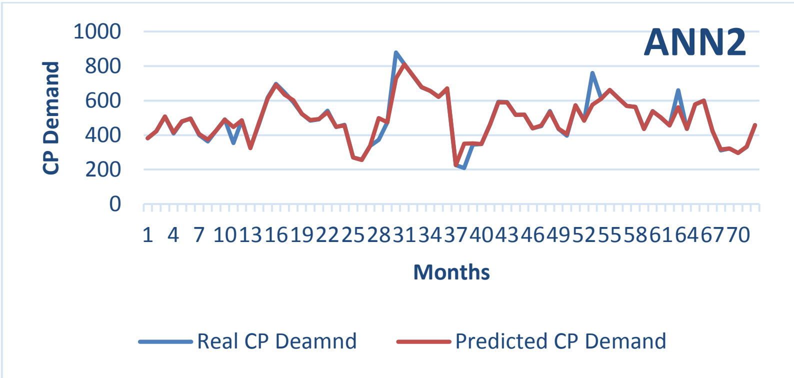

Figure 3. Predicted and Real CP demand of ANN2

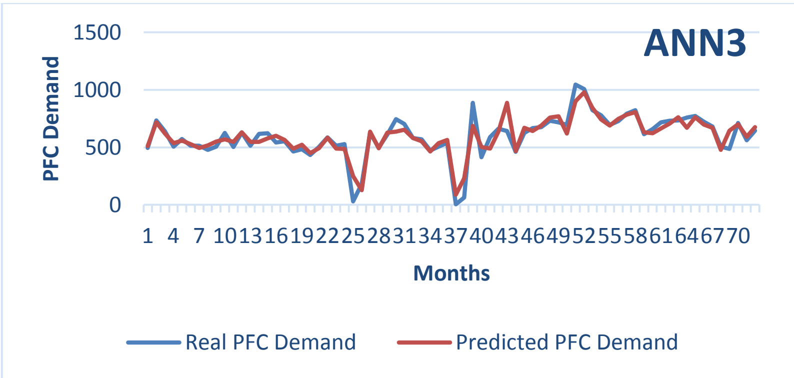

Figure 4. Predicted and Real PFC demand of ANN3

Acknowledgements

The authors gratefully acknowledge all the staff of the Regional Transfusion Blood Center (CRTS) of Fez city for providing demands of blood components, especially the director Dr. Abderrahim Benyasrhi, Pr. Soumia Zizi and Dr. Souad Salhi. The authors are also thankful to the University Hospital Hassan II of Fez city (CHU Hassan II) for affording data on all the variables using in the models.

7. REFERENCES

[1] Oscar, S. Silva Filho, W. Cezarino, and Giselle R. S. 2012. A Decision-making Tool for Demand Forecasting of Blood Components. *Proceedings of the 14th IFAC Symposium on Information Control Problems in Manufacturing Bucharest, Romania*, 23-25, 1499-1504.

[2] Lang, J. C. 2010. Blood bank inventory control with transshipments and substitutions, production and inventory mnagement with substitutions. *Lecture Notes in Economics and Mathemat. Systems*, 636, 205-226. DOI= http://dx.doi.org/10.1007/978-3-642-04247-8_8.

[3] Karaesmen, I. Z., Scheller–Wolf, A., and Deniz, B. 2011. Planning production and inventories in the extended enterprise. *International Series in Operations Research & Management Science*, 151, 393-436. DOI= http://dx.doi.org/10.1007/978-1-4419-6485-4.

[4] Sima, M.F., and Elena, A.K. 2016. Reducing uncertainty in demand for blood. *Operations Research for Health Care 9*, 16–28. DOI= http://dx.doi.org/10.1016/j.orhc.2016.02.002.

[5] Uthayakumar, R., and Priyan, S. 2013. Pharmaceutical supply chain and inventory management strategies: Optimization for a pharmaceutical company and a hospital. *Oper. Res. Health Care 2*, 52–64. DOI= http://dx.doi.org/10.1016/j.orhc.2013.08.001.

[6] Coulibaly, P., Anctil, F., and Bobée, B. 2000. Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. Journal of Hydrology 230, 244-257. DOI= http://dx.doi.org/10.1016/S0022-1694(00)00214-6.

[7] Chopra, S., and Meindl, P. 2013. Supply Chain Management: Strategy, Planning, and Operation. *Pearson Education, Inc., Upper Saddle River*, NJ. DOI= http://dx.doi.org/10.1007/978-3-8349-9320-5_22.

[8] Yen, H. 1975. *Inventory Management for a Perishable Product Multi-Echelon System*. PhD Dissertation. Department of Industrial Engineering and Management Science. Northwestern University, Evanston, IL.

[9] Brandeau, M.L., Sainfort, F., and Pierskalla, W.P.2004. Supply chain management of blood banks. Operations Research and Health Care: A Handbook of Methods and Applications, Kluwer Academic Publishers, 103–145. DOI= http://dx.doi.org/10.1007/b106574.

[10] Frankfurter, G.M., Kendall, K.E., and Pegels, C.C. 1974. Management control of blood through short-term supply-demand forecast system. Management Science 21 (4) 444–452. DOI= http://dx.doi.org/10.1287/mnsc.21.4.444.

[11] Muluneh, A. 2014. *Control Chart Pattern Recognition for Multivariate Autocorrelated Processes Using Artificial Neural Network*, Thesis, 1-96.

[12] Guh, R.S., and Tannock, J.D.T. 1999. Recognition of control chart concurrent patterns using a neural network approach, International Journal of Production Research, 37:8, 1743-1765. DOI= http://dx.doi.org/10.1080/002075499190987.

[13] Perry, B.M., Spoerre, K.J. and Velasco, T. 2001. Control chart pattern recognition using back propagation artificial neural networks. International Journal of Production Research, 39:15, 3399-3418. DOI= http://dx.doi.org/10.1080/00207540110061616.

[14] Yu, S., Zhu, K., and Zhang, X., 2012. Energy demand projection of China using a path-coefficient 646 analysis and PSO–GA approach. *Energy Conversion and Management 53*, 142-153. DOI= http://dx.doi.org/10.1016/j.enconman.2011.08.015.

[15] Çelik Ö, Teke A, and Yildirim HB. 2016. The optimized artificial neural network model with Levenberg-Marquardt algorithm for global solar radiation estimation in Eastern Mediterranean Region of Turkey. Journal of Cleaner Production. DOI= http://dx.doi.org/10.1016/j.jclepro.2015.12.082.

[16] Kalogirou, S.A., 2000. Applications of artificial neural-networks for energy systems. *Applied Energy 67*, 17-35.

[17] Smith, G. N. 1986. Probability and statistics in civil engineering: An introduction. London: Collins. DOI= http://dx.doi.org/10.1155/2013/268064.

[18] Blake, J.T., and Hardy, M. A generic modelling framework to evaluate network blood management policies: The Canadian Blood Services experience, *Oper. Res. Health Care 3* (2014) 116–128.

[19] Belien, J., and Force, H. Supply chain management of blood products: A literature review, European J. Oper. Res. 217 (2012) 1–16. DOI= http://dx.doi.org/10.1016/j.promfg.2015.11.007.

[20] Bosnes, V., Aldrin, M., and Heier, H.E. Predicting blood donor arrival, Transfusion 45 (2) (2005) 162–170. DOI= http://dx.doi.org/10.1111/j.1537-2995.2004.04167.x.