Asian Journal of Information Technology 15 (22): 4475-4481, 2016
ISSN: 1682-3915
© Medwell Journals, 2016

A Location-Based Framework for Mobile Blood Donation and Consumption Assessment Using Big Data Analytics

Sherin Moussa
Department of Information Systems, Faculty of Computer and Information Sciences,
Ain Shams University, 11566 Cairo, Egypt

Abstract: Recently, the advancements in communication technology accompanied by the widespread availability of mobile devices have greatly evolved mobile health care applications. Blood donation systems are one of the crucial management systems in health sector, where instant responses to immediate needs for specific blood group in case of emergencies is vital. Thus, the fulfillment of blood demands in the right time from the nearest blood banks draws a necessity to efficiently direct donors to the right location to donate. In this study, we propose a Location-based Analyzer for Mobile Blood Donation Assessment (LAMBDA) framework. The proposed framework uses large-scale time series regression analysis techniques to analyze blood demands and donations matching data initiated from a mobile application and forecast blood shortages and wastages per blood group for a specific location. Accordingly, donors can be directed to the nearest location having shortage of their blood group. Hence, blood wastage rates are improved dramatically.

Key words: Big data analytics, large scale time series regression analysis, prediction, blood donation, blood donor app, location-based services, mobile computing

INTRODUCTION

Considering the life-threatening nature of Blood Donation and Transfusion (BDT) services and strict storage, safe blood availability in blood banks and health organizations is critical and vital in a reliable healthcare system. Blood banks usually suffer frequent shortages of certain blood groups at some locations. Hence, social networks have recently served as a quick channel for advertisements, looking for healthy individuals to donate blood for patients who urgently require blood transfusionat emergencies (Qin, 2014). Whereas, some other blood donation centers may have excess of the same required blood group which it would eventually be wasted for expiry reasons. This is due to the fact that many blood banks work in isolation, having no integration with other health organizations that would dramatically affect the quality of BDT services. Moreover, BDT processes usually consume long time and effort from demanders, medical staff and donors. One main reason is the lack of reliable information system that allows direct coordination among these parties in order to minimize time and effort required for such processes. Recently, some applications on mobile phones have been developed to initiate social communities related to BDT services in many territories.

An efficient blood bank management system should provide sufficient quantity of qualified blood to demanders at the right time within their nearest location. The rapidly increase of huge amounts of data in the blood bank sector obliges to efficiently study these data using big data analytic techniques to perform accurate forecasting of future demands and donations, to analyze inventory and consumption patterns and to monitor performance metrics. Thus, shortage and wastage of blood can be handled, leading to self-sufficiency in blood requirement in addition to an effective planning for blood donation campaigns in the right location and timing (Rakthanmanon *et al*., 2013; Schreiber *et al*., 2003).

In this study, we propose a Location-based Analyzer for Mobile Blood Donation Assessment (LAMBDA) framework based on the leading-edge information technologies of big data analytics and mobile computing. The proposed framework server-side uses large-scale time series regression analysis to analyze urgent blood demands and their associated donations data per location that are initiated from our developed mobile application, as well as the related blood consumption and wastage patterns per blood group. Time-series data analysis has been used in diverse applications to discover valuable patterns extracted from complex datasets. Huge datasets make standard data mining solutions almost impossible, considering big data analytics instead (Rakthanmanon *et al*., 2013). Whilst the mobile-side provides a direct location-based communication channel between blood demanders, donors and blood-related organizations for real-time aid.

The rest of the study is organized as follows: section two highlights the main related studies that have been conducted in the field of mobile location-based blood donation systems and large-scale time series analysis. Section three presents our proposed framework. Section four discusses the experiments and their associated results to evaluate the proposed framework. Section five concludes our study and future work.

Literature review: By the rapid evolution of technology and medical advances, many studies have addressed BDT services. Belien and Force (2012) an extensive study has been conducted on the inventory and supply chain management of blood products, distinguishing the different applied solutions to blood related problems. Another study about the automation of blood donation classification and notification using machine learning techniques was presented in Chinnaswamy *et al*. (2015) for aiding and enhancing donation and distribution. Delen *et al*. (2011), a web-based Blood Reserve Availability Assessment, Tracking and Inventory Management System (BRAMS) was introduced where data mining and GIS-based analytics were combined to support decision making in a proactive blood supply chain management. Another decision support system named (SIABAS) was implemented in Li *et al*. (2008) based on a hybrid paradigm of rule-based expert systems, data-driven statistical analyses and Oracle Data Mining toolkit (ODM) for donor screening, donation evaluation, sample testing, blood processing and product dissemination. Another prediction model was designed in Drackley *et al*. (2012) to forecast Ontario's blood supply and demand. The donation histories of donors from five geographically diverse blood centers were studied in Schreiber *et al*. (2003) in order to predict the number of additional donations that could be collected to increase blood availability by changing donation patterns and increasing the donor return rate. Several models were proposed in Ding *et al*. (2015), Islam *et al*. (2013), Simmhan and Noor (2013), Testik *et al*. (2012), Boonyanusith and Jittamai, Sharma and Gupta (2012), Zabihi *et al*. (2011), Santhanam and Sundaram (2010) and Bosnes *et al*. (2005) to predict the future behavior of blood donation of blood donors, discover their arrival patterns and to study the response to scheduled donation appointments. Whereas researchers in Khalilinezhad have built a classification model to predict healthy donors without visiting the doctor or doing laboratory tests. Different predictive classification approaches were used in these works including k-means clustering, J48, CART, Weka, neural networks, decision tree, naïve bayesian, SVM and fuzzy sequential pattern mining approaches.

A set of mobile applications have been developed as well to link donors in order to improve BDT service quality, so that the appropriate donor can be reached just on time. Jenipha and Backiyalakshmi (2014), a web-based system using SMSs and a Geo-location RVD Scoring Algorithm was proposed while researchers in Singh *et al*. (2007) developed mobile applications connected to a cloud-based systems. Different mobile applicationswere developed by Rahman *et al*. (2011) based on SMSs in a specific format. Gupta *et al*. (2015) and Premasudha *et al*. (2009), others were developed to find spatial distribution of blood donors where GIS was applied to locate the blood donors of the required blood group during emergency. Whereas in Jiang *et al*. (2005) an RFID-based blood management system on smart devices was introduced to ensure a reliable and credible process of blood donor identification and to increase the operation management efficiency.

On the other handwith the recent evolution of mobile and cloud computing, big data and their vast applications, many research efforts have been emerged concerning the analysis of time series data in essence of big data. A brief introduction was presented in Qin (2014), concluding the main tasks in time series data mining research with regular sampling periods. Whilst in Aghabozorgi *et al*. (2015), Wang *et al*. (2015), Ding *et al*. (2015) and Simmhan and Noor (2013), large-scale time-series clustering algorithms were investigated, having some used raw time-series and others tried to use reduction methods or the Auto-Regressive Integrated Moving Average (ARIMA) model before clustering of time-series data. Wickham (2011) an R package named speed-wise plyr was developed based on the split-apply-combine strategy for data analysis where a large process was broken up into manageable pieces, operated on each piece independently and then all the pieces were put back together. Chen (2014) a high-order fuzzy time series model was proposed based on entropy-based partitioning and adaptive expectation model for time series forecasting. Another studies were held in Jirkovsky *et al*. (2014) and Mirko and Kantelhardt

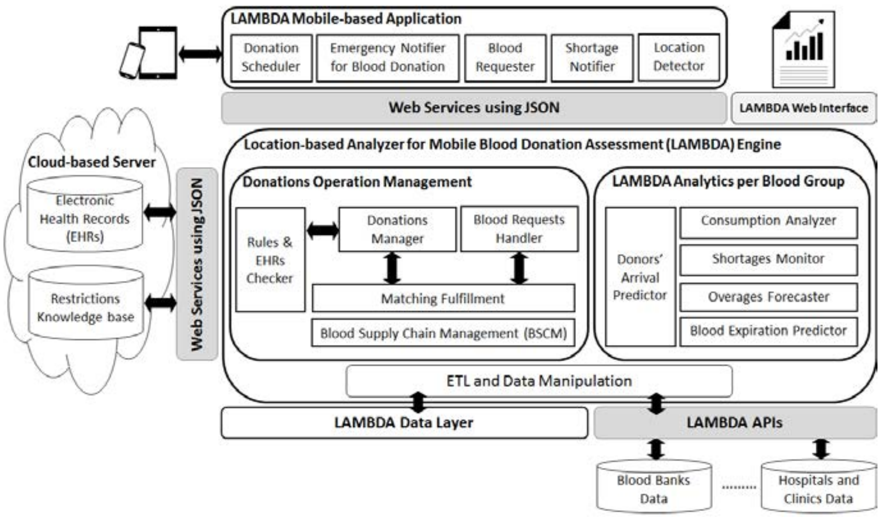
LAMBDA Mobile-based Application
Donation Scheduler
Emergency Notifier for Blood Donation
Blood Requester
Shortage Notifier
Location Detector
Web Services using JSON
LAMBDA Web Interface
Location-based Analyzer for Mobile Blood Donation Assessment (LAMBDA) Engine
Cloud-based Server
Electronic Health Records (EHRs)
Restrictions Knowledge base
Web Services using JSON
Donations Operation Management
Rules & EHRs Checker
Donations Manager
Blood Requests Handler
Matching Fulfillment
Blood Supply Chain Management (BSCM)
LAMBDA Analytics per Blood Group
Donors' Arrival Predictor
Consumption Analyzer
Shortages Monitor
Overages Forecaster
Blood Expiration Predictor
ETL and Data Manipulation
LAMBDA Data Layer
LAMBDA APIs
Blood Banks Data
....
Hospitals and Clinics Data

Fig. 1: Divided time series regression analysis flowchart

(2013) to examine the utilization of Hadoop and Hadoop.TS library on time series data, following the Partitioning design pattern of Map-Reduce. Whereas in Qin (2014), a parallel time series forecasting approach was introduced based on ARIMA model where it parallelized computation by splitting the data into n chunks, getting the results from previous iteration as the initial values for the current iteration. Another Least Squares-Support Vector Machine method (LS-SVM) with Kernel Principal Component Analysis (KPCA) and Particle Swarm Optimization (PSO) was introduced in Chen *et al*. (2010) for time series power load forecasting.

Although, the previous studies considered several aspects of the BDT cycle, however each studyfocused on one aspect only, ignoring the effectiveness of the rest of facets on the overall BDT service quality. In addition, minimal research has directed big data analytics toanalyze the huge datasets of blood banks. One of the main reasons for the uniqueness of the proposed research is that it applies large-scale time series regression to blood banks sector, linking blood demands with donations to forecast consumption rates, estimated shortages, overages and wastages per blood group over regions.

MATERIALS AND METHODS

LAMBDA framework: In order to address an efficiently-integrated framework for a qualified BDT service using big data analytics, we introduce our Location-based Analyzer for Mobile Blood Donation Assessment (LAMBDA) framework. Figure 1 shows the system architecture of our proposed framework. It consists of four main components as follows.

LAMBDA cloud-based server and APIs: A cloud-based knowledgebase for the restrictions of donations in addition to a cloud-based Electronic Health Records (EHRs) have been integrated to the proposed framework to perform the first screening and filtering of donors, thus minimizing the donated blood from non-eligible donors. Standardized APIs have been developed to integrate different blood banks, blood donation centers and health organizations for better service area coverage. Table 1 presents s sample for the restrictions of donations included in the knowledge base.

LAMBDA mobile-based interface: LAMBDA mobile applicationis developed to allow donors starting the blood donation process through the donations scheduler modulewith minimal effort and time at the right nearest location where shortage of their blood group has been detected. Blood demanders on the other hand, can broadcast instant blood requests through blood requester module. Donors would then receive notifications about such urgent blood demands and the detected stock shortages through the emergency notifier and the shortage notifiermodules respectivelyfor their own blood group sorted by the nearest location. Location detector module serves all the former modules to determine the current user's location.

LAMBDA server-side engine: This is the main core of our proposed framework where the data processing and blood-related analytical studies are performed. It includes three main modules as detailed below in the following sub-sections.

Table 2: Sample for the donations' restrictions in the knowledgebase

Description of restricted diagnosis	**Description of restricted diagnosis**
Last donation≤8 weeks	Fever
Age≤17	Productive cough
Weight≤110 lbs	AIDS
Female height≤5'6"	Hepatitis viruses
Male height≤5'	Pregnancy
80/50≥Blood pressure≥180/100	Recent childbirth
100≤Pulses≤50	Low level of iron-hematocrit
Hemoglobin<12.5 g dL^{-1}	Blood cancer
on Antibiotics	Leukemia
Flu	Lymphoma
Non-melanoma skin cancer	Malaria≤1 year of treatment
Bleeding problems	Taking any "blood thinner"
Taking anticoagulants	Heart attack≤6 months
Episode of angina≤6 months	Had surgery/angioplasty≤6 months

Breast/ Brain/Prostate/Lung cancer ≤5 years after diagnosis or the last surgery/chemotherapy/radiation treatment

ETL and data manipulation: This is a mediator layer which receives blood data collected from hospitals, clinics, blood donation centers and blood banks integrated through LAMBDA APIs. A customized Extract, Transform, Load (ETL) approach is used to unify the data formats extracted from the different sources into our framework. Data manipulation techniques are then applied for data cleansing. The resultant transformed data are then loaded into LAMDBA data layer.

Donations operation management: It handles the Online Transactional Processing (OLTP) capabilities of LAMBDA framework. Donation appointments are received through the donations manager sub-module. potential donors are directly checked by the Rules and EHRs Checker sub-module for their eligibility through the integratedcloud-based restrictions knowledgebase and EHRs. Whereas blood requests handler receives all emergency demands, identifies the required blood group and location and interacts with the Matching Fulfillment (checking upcoming donations), BSCM (checking available valid stock) and emergency notifier (to notify donors) sub-modules in order to satisfy demands. BSCM sub-module aggregates blood data from data sources through the ETL and data manipulation module. Location-based instantaneous updates are applied to expirations and quantities per blood group.

LAMBDA analytics per blood group: A large-scale time series regression analysis approach is appliedas a big data analytic technique to evaluate the blood supply chain status, to identify the blood posture readiness for emergencies and to analyze inventory and consumption patterns within regions. Donors' arrival predictorand consumption analyzer sub-modules are first triggered in order for the rest of the analytics to function simultaneously. Time series regression analysis depends on the autocorrelation between time events. This allows LAMBDA framework toforecast imbalances and wastages in the inventory positions of different blood groups within regions in terms of shortages or overages. Regression and ARIMA approaches are the most well-known methods to time series analysis. However, regression analysis is superior compared to ARIMA, as regression produces better models and is more powerful and flexible. Besides, regression uses a "closed" computational algorithm that is essentially guaranteed to yield results if at all possible while ARIMA uses iterative algorithms that often fail to reach a solution (Zabihi *et al*., 2011). We use the Divided Regression Analysis approach Jun *et al*. (2015) where the big time series data, representing the population dataset for blood donations and demands, are divided into n sub datasets with small size based on blood group and region, each representing asample. Figure 2 shows the flowchart of divided time series regression analysis applied in LAMBDA framework. The multiple linear regression model for the population is represented as follows O'-Connell and Koehler:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \epsilon \quad (1)$$

Where:

Y = Adependent variable

$X_{1,} X_{2 \dots} X_k$ = Independent variables

$\beta_, \beta \dots \beta_k$ = Regression parameters

$\in$ = The error of the model

In order to estimate the regression parameter vector $B = (\beta_{0,} \beta_1 \dots \beta_k)$ for the whole population, we compute an estimate parameter vector = $\hat{B} = \hat{B}_1 \hat{B}_2 \dots \hat{B}_n$ for each of the n sample sub-datasets. Thus, it ends up with $\hat{B} = \hat{B}_1 \hat{B}_2 \dots \hat{B}_n$ for n sub-datasets.The n estimated parameter vectors are then combined using the mean value combine

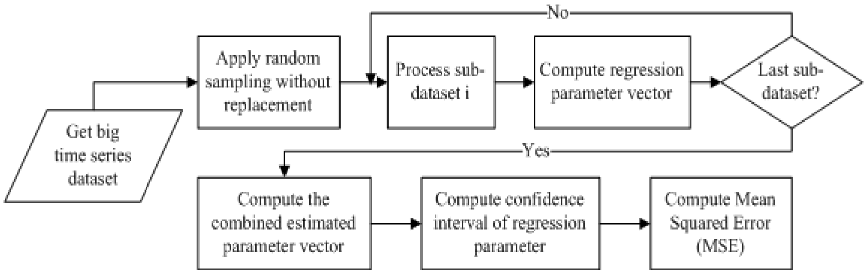
Get big time series dataset
Apply random sampling without replacement
Process sub-dataset i
Compute regression parameter vector
Last sub-dataset?
No
Yes
Compute the combined estimated parameter vector
Compute confidence interval of regression parameter
Compute Mean Squared Error (MSE)

Fig. 2: Divided time series regression analysis flowchart

function to estimate B for the original population through $\hat{B}_c$ (Jun *et al*., 2015), using the Mean Squared Error (MSE) to verify results as follow:

$$\hat{B}_c = y_n^1 \sum_{i=1}^{n} \hat{B} \tag{2}$$

$$MSE = 1/n\Sigma_1 (i=1)^{\uparrow} n \left|(Y)\right|_i - Y_i)^2 \tag{3}$$

LAMBDA web-based Interface: This is the analytical user interface for reports and data visualization where all the resultant analysis and forecasting are displayed, together with the associated appropriate graphical representations. In addition, medical staff can use LAMBDA web to modify the EHRs of donors proven to be eligible/non-eligible for donation.

RESULTS AND DISCUSSION

A dataset of 622,114 donors and 515,229 demanders with 7,155,904 donation appointments and blood requests collected over 22 months from different 11 regions was used to evaluate our proposed LAMBDA framework. This huge dataset was split into 88 sub-datasets (i.e., 8 blood groups; O-,O +, A-, A+, B-, B+, AB-, AB+, spread among 11 regions). This was integrated with a cloud-based EHRs of 336,216 records. Experiments have investigated the precision and recall evaluation metrics for the prediction of donors' arrival, blood wastage rates due to expiry reasons in addition to the shortage and overages rates of the different blood groups within the different 11 regions. As shown in Fig. 3 and 4 for the shortages rates prediction, the precision value increases by time, starting by 57% at the second month for the 8 blood groups in the 11 regions, till it reaches 92% at the twenty-second month. Whereas recall value started by

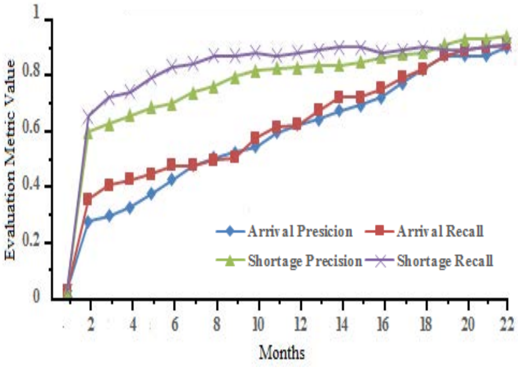

Fig. 3: Precision and recall values of donor's arrival and shortage rates prediction at LAMBDA

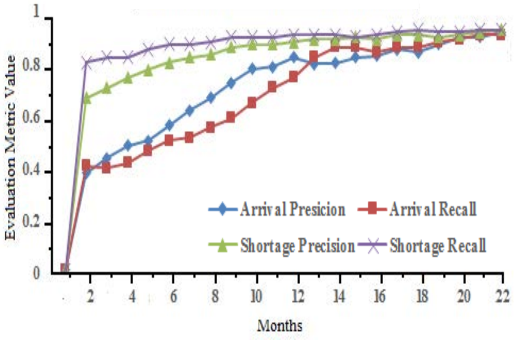

Fig. 4: Precision and recall values of overage and wastage rate prediction at LAMBDA

same period. Regarding the donor's arrival rate prediction, precision values varied from 25% till 88%, while recall percentages ranged from 33% till 89%. Thus, the average precision and recall values of our proposed framework regarding the shortages rate prediction are 78 and 83%, besides 59 and 62% for the donor's arrival rate prediction respectively. Figure 4 shows the precision values of the overage rate prediction, ranging from 38-94% during the 22 months while the recall values ranged between 41 and 93%. It also presents the precision values of the wastage rate prediction which recorded percentages varied from 68% till 95% whilst its associated recall percentages varied from 82% till 95%. Hence, our proposed framework achieves an average of 74 and 71% for the overage rates, as well as 87 and 91% for the wastage rate's precision and recall evaluation metrics, respectively.

CONCLUSION

In this study, we propose a Location-based Analyzer for Mobile Blood Donation Assessment (LAMBDA) framework using a large-scale time series regression approach for big data analytics. LAMBDA framework can analyze consumption and wastage patterns based on the blood donations and emergency requests in order to forecast blood shortages and overages per blood group for a specific location. Accordingly, donors can be directed to the nearest location having shortage of their blood group. The results of our experiments show that our proposed framework achieves an average of 74 and 71% for the precision and recall for the overages rate prediction, 87% and 91% for the wastage rates, whilst the average precision and recall values for the donor's arrival rate prediction are 59 and 62%, as well as the related values to the shortages rate prediction are 78% and 83% respectively.

RECOMMENDATIONS

The future work includes the enhancement of the prediction results of the framework, the investigation of other big data analytic technique to study the blood sector data in addition to adding new facilities to the framework to serve different levels of granularity.

REFERENCES

Aghabozorgi, S., A.S. Shirkhorshidi and T.Y. Wah, 2015. Time-series clustering: A decade review. Inf. Syst., 53: 16-38.

Belien, J. and H. Force, 2012. Supply chain management of blood products: A literature review. Eur. J. Oper. Res., 217: 1-16.

Bosnes, V., M. Aldrin and H.E. Heier, 2005. Predicting blood donor arrival. Transfusion, 45: 162-170.

Chen, M.Y., 2014. A high-order fuzzy time series forecasting model for internet stock trading. Future Generation Comput. Syst., 37: 461-467.

Chen, Q., X. Chen and Y. Wu, 2010. Optimization algorithm with Kernel PCA to support vector machines for time series prediction. J. Comput., 5: 380-380.

Chinnaswamy, A., G. Gopalakrishnan, K.K. Pan-dala and S.N. Venkata, 2015. A study on automation of blood donor classification and notification techniques. Intel. J. Appl. Eng. Res., 10: 18503-18514.

Delen, D., M. Erraguntla, R.J. Mayer and C.N. Wu, 2011. Better management of blood supply-chain with GIS-based analytics. Ann. Operations Res., 185: 181-193.

Ding, R., Q. Wang, Y. Dang, Q. Fu and H. Zhang *et al*., 2015. Yading: Fast clustering of large-scale time series data. Proc. VLDB Endowment, 8: 473-484.

Drackley, A., K.B. Newbold, A. Paez and N. Heddle, 2012. Forecasting Ontario's blood supply and demand. Transfusion, 52: 366-374.

Gupta, N., R. Gawande and N. Thengadi, 2015. MBB: A life saving application. Intl. J. Res. Emerging Sci. Technol., 2: 2349-7610.

Islam, A.S., N. Ahmed, K. Hasan and M. Jubayer, 2013. Health: Blood donation service in Bangladesh. Proceedings of the 2013 International Conference on Informatics, Electronics and Vision (ICIEV), May 17-18, 2013, IEEE, New York, USA., pp: 1-6.

Jenipha, T.H. and R. Backiyalakshmi, 2014. Android blood donor life saving application in cloud computing. Am. J. Eng. Res., 3: 105-108.

Jiang, M., B. Xing, Z. Sun, P. Fu and H. Chen et al., 2005. A dynamic blood information management system based on RFID. Proceedings of the IEEE Conference on Engineering Medicine Biology Society, September 1-4, 2005, IEEE, New York, USA., pp: 546-549.

Jirkovsky, V., M. Obitko, P. Novak and P. Kadera, 2014. Big data analysis for sensor time-series in automation. Proceedings of the 2014 IEEE Conference on Emerging Technology and Factory Automation (ETFA), September 16-19, 2014, IEEE, New York, USA., ISBN: 978-1-4799-4845-1, pp: 1-8.

Jun, S., S.J. Lee and J.B. Ryu, 2015. A divided regression analysis for big data. Stat., 9: 21-32.

Li, B.N., M.C. Dong and S. Chao, 2008. On decision making support in blood bank information systems. Expert Syst. Appl., 34: 1522-1532.

Mirko, K. and J.W. Kantelhardt, 2013. Hadoop. TS: large-scale time-series processing. Intl. J. Comput. Appl., Vol.74,

Premasudha, B.G., S. Swamy and B.S. Adiga, 2009. An application to find spatial distribution of blood donors from blood bank information system. Intkl. J. Inf. Technol., 2: 401-403.

Qin, S.J., 2014. Process data analytics in the era of big data. AICHE. J., 60: 3092-3100.

Rahman, M.S., K.A. Akter, S. Hossain A. Basak and S.I. Ahmed, 2011. Smart blood query: A novel mobile phone based privacy-aware blood donor recruitment and management system for developing regions. Proceedings of the 2011 IEEE Workshops of International Conference on Advanced Information Networking and Applications (WAINA), March 22-25, 2011, IEEE, New York, USA., pp: 544-548.

Rakthanmanon, T., B. Campana, A. Mueen, G. Ba-tista and B. Westover et al., 2013. Addressing big data time series: Mining trillions of time series subsequences under dynamic time warping. ACM. Trans. Knowl. Discovery Data, 7: 1-10.

Santhanam, T. and S. Sundaram, 2010. Application of CART algorithm in blood donors classification. J. Comput. Sci., 6: 548-552.

Schreiber, G.B., A.M. Sanchez, S.A. Glynn and D.J. Wright, 2003. Increasing blood availability by changing donation patterns. Transfusion, 43: 591-597.

Sharma, A.K. and P.C. Gupta, 2012. Predicting the Number of Blood Donors through their Age and Blood Group by using Data Mining Tool. Int. J. Commun. Comput. Technol., 1: 1-6.

Simmhan, Y. and M.U. Noor, 2013. Scalable prediction of energy consumption using incremental time series clustering. Proceedings of the IEEE International Conference on Big Data, October 6-9, 2013, IEEE, New York, ISBN: 978-1-4799-1293-3, pp: 29-36.

Singh, R., P. Bhargava and S. Kain, 2007. Smart phones to the rescue: the virtual blood bank project. IEEE. Pervasive Comput., 4: 4-89.

Testik, M.C., B.Y. Ozkaya, S. Aksu and I.O. Ozcebe, 2012. Discovering blood donor arrival patterns using data mining: A method to investigate service quality at blood centers. J. Med. Syst., 36: 579-594.

Wang, X., F. Yu, H. Zhang, S. Liu and J. Wang, 2015. Large scale time series clustering based on fuzzy granulation and collaboration. Intl. J. Intell. Syst., 30: 763-780.

Wickham, H., 2011. The split-apply-combine strategy for data analysis. J. Stat. Software, 40: 1-29.

Zabihi, F., M. Ramezan, M.M. Pedram and A. Me-mariani, 2011. Rule extraction for blood donators with fuzzy sequential pattern mining. J. Math. Comput. Sci., 2: 37-40.