American International Journal of Multidisciplinary Scientific Research
Vol. 10, No. 1; 2021
ISSN 2638-1249 E-ISSN 2638-1273
Published by CRIBFB, USA

DELVING INTO BLOOD TRANSFUSIONS DATA THROUGH DATA MINING: A STUDY OF MAIZBHANDARI SHAH EMDADIA BLOOD DONORS GROUP TO SELECT VOLUNTEER BLOOD DONORS EFFICIENTLY

Syed Irfanul Hoque
Managing Trustee of Darul Irfan Research Institute (DIRI) and **Nayeb Sajjadah Nasheen** of
Maizbhandar Darbar Sharif, Fatikchari, Chattogram, Bangladesh
E-mail: tasauf.darulirfan@gmail.com

Md. Minhazul Abedin
Student of Department of Computer Science and Telecommunication Engineering**,**
Noakhali Science and Technology University, Bangladesh
Associate Member of DIRI, Chattogram, Bangladesh
E-mail: tasauf.darulirfan@gmail.com

Mohammad Sohel Chowdhury
Student of Department of Computer Science and Telecommunication Engineering**,**
Noakhali Science and Technology University, Bangladesh
Associate Member of DIRI, Chattogram, Bangladesh
E-mail: tasauf.darulirfan@gmail.com

ABSTRACT

The demand for blood transfusion is rising gradually. Therefore, volunteer blood donors are needed to save the lives of patients. There are lot of volunteer donors in every society. But, the main problem people face is finding such donors at the right time. To locate potential blood donors, we collected data from the Maizbhandari Shah Emdadia blood donors' group, which consists of 700 active volunteer blood donors. This study aims to choose potential volunteer donors efficiently at the emergency time from the Maizbhandari Shah Emdadia blood donors' group based on their past data. We have developed two models, namely descriptive and predictive models, using data mining techniques. The descriptive model analyzes data patterns and explores the donors' behaviour. A data mining clustering algorithm was used to develop the descriptive model. The underlying factors of donors' intention to donate blood were identified and evaluated using this descriptive model. These factors were then utilized to develop the predictive model, which in turn assists to predict whether a donor will donate blood or not during an emergency. The findings of these two models will assist the clinical experts in locating potential volunteer blood donors within the shortest period and thus save valuable lives.

Keywords: Data Mining, Maizbhandar Blood Donation, Clustering, Classification, Prediction Model.

INTRODUCTION

Blood donation is the best goodwill of a person. It plays a significant role in saving lives. Every day, we see a massive demand for blood all across the world. The need for blood is increasing as a result of surgical treatment for various diseases (Gulinac, 2020; Serteva, 2019), accidents, delivery cases, kidney diseases (Velikova et al., 2018), thalassemia, anemia, and extreme health issues such as organ transplants, Leukemia, and bone marrow cancer. All of these urgent scenarios result in sudden high blood loss, necessitating an emergency blood supply, which if not provided can result in death.

Each year, 118.4 million units of blood are collected worldwide, according to the World Health Organization (WHO). Nearly half of them are gathered in high-income countries. Blood donation rates in high-income countries, upper-middle-income countries, lower-middle-income countries, and low-income countries are 31.5, 15.9, 6.8, and 5.0 donations per 1000 people, respectively. Especially, South-East Asia faces significant difficulties in the shortage of blood. ("Blood safety and availability," 2020)

There are three types of blood donors: volunteer, family/friends of the patient, and paid donors. According to the WHO, 79 countries get over 90% of their blood supply from unpaid volunteers. Over half of the blood supply in 56 countries comes from family/replacement or paid donors ("Blood safety and availability," 2020). These findings show that many high-income countries meet their blood demand through volunteer donors. Since there is a shortage of blood in developing countries (World Health Organization, 2010), volunteer blood donation is the best option to reduce this shortage. Again, Bangladesh, one of the developing countries, receives just 31% of its blood from volunteers. This figure is quite low when compared to other South-East Asian countries like Thailand, India, and Sri Lanka, where the figure can reach as high as 95 percent ("Bangladesh is still to meet the demand of safe blood supply," 2017). To increase volunteer donors, the Bangladesh government set a target to guarantee 100% blood collection through voluntary donations by 2022 (Al-Masum Molla, 2017). Hence, there are many public or private organizations working together to reduce the shortage of blood. Hazrat Maulana Shah Sufi Syed Emdadul Hoque Maizbhandari (M.J.A) formed one of the organizations, the "Maizbhandari Shah Emdadia blood donors' group," to serve humanity by donating blood in particular and required situations. The members of this organization encourage and organize themselves and their siblings or relatives in various professions to prepare and maintain a list of blood donors, including blood group identification, address, and mobile number, which is working as a data centre. Later, blood recipients can easily collect blood by collecting information about blood donors from the data centre.

Problem Statement

Generally, there is a shortage of blood in Bangladesh. According to the Directorate General of Health Services (DGHS), Bangladesh requires 1-1.2 million bags of blood each year. However, in 2018, 7,73,383 bags of blood were donated (Tithila, 2019). To reduce this shortage voluntary blood donation may be the best option. Lately, there is significant increase of 2.37 million voluntary unpaid donors from 2013 to 2018 across South-East Asia ("Blood safety and availability," 2020). Although the number of donors is increasing, there is still a problem finding willing donors at the emergency time in the shortest period. The study of Sahariar Hasan Jiisun et al. (2019) showed that about 48% of blood seekers in Bangladesh require 19-24 hours to manage each bag of blood. This time gap can cause the loss of life of a patient. An efficient selection of the most probable donors from a group of donors is required to reduce this time gap. Therefore, we have decided to conduct our research on the Maizbhandari Shah Emdadia blood donors' group's data to solve this issue.

Objectives

The objectives are as follows:

- To efficiently select potential volunteer blood donors from a group when patients require blood in emergency time.
- To minimize the time gap between demand and supply of blood donation.

Research Questions

This paper has tried to answer the following questions:

- How to select the most probable volunteer donor at the emergency time from a group?
- How to minimize the time gap between demand and supply of blood donation?

LITERATURE REVIEW

Data Mining in Blood Donation

Data mining is a method of extracting information from huge volumes of data. The goal of data mining is to find out valuable and relevant information. Data mining is utilized in a variety of applications, including financial analysis (Lakshmi & Kumar, 2018), education, AI systems (Adetayo Olaniyi & Olufunto Adedotun, 2018), and medical healthcare. Because of developments in medical informatics and information technology, the blood donor information system has rapidly risen in size. The growing demand for high-quality blood donations requires maximizing the potential of stored data, not just clinical data, to enhance blood donor behavior. The blood donation sector could benefit significantly from data mining. It can be a useful tool for examining data collected by blood donation organizations via their information systems. (Bhardwaj, 2012).

Agarwal et al. (2019) suggested a blood transfusion system using various data mining algorithms such as Random Forest, Decision Tree, and Logistic Regression. They found that the Random Forest algorithm shows the highest efficiency in predicting whether a person will donate blood or not, based on the past data of donation. Hence, they used the Random Forest algorithm to develop a model for predicting the nature of a donor.

KIRCI et al. (2020) tried to estimate if a possible donor will provide blood donation again using Logistic Regression and Naive Bayes data mining algorithms. They compared these two algorithms and found that the Logistic regression outperformed the Naive Bayes classifier. Alkahtani et al. (2019) had been collected data over two years (2017-2018) from the Saudi blood bank. They applied Random Forest, Support Vector Classifier (SVC), and Logistic Regression algorithms on the collected data for classifying blood donors as a return and non-return donors. In addition, a time series analysis was applied to explore the seasonal variations in blood donation. The result showed that the blood donation rate drops in June and September due to two religious periods: Fasting and Hajj. Another study (Alajrami, 2019) was conducted for predicting whether a person is going to donate blood or not using an Artificial Neural Network. They got the highest prediction accuracy (99.31%) than previous studies in the UCI Machine Learning repository (Yeh et al., 2009) blood transfusion dataset. Ashoori et al. (2013) used the clustering method to identify blood donors' behaviour. After analyzing donors' behaviour they found that most donors were men and had O positive blood group.

Web-based Blood Donation System

Recently, web-based blood donation management systems have become very popular. A web-based blood donation management system allows anyone who wants to donate blood to help others in need. It also allows hospitals to keep track of and preserve information for anyone who wants to contact them and provides a centralized blood bank database (Hashim, 2014).

Fahad et al. (2019) developed a web application for android mobiles. This application links all the donors and provides a list of blood banks in any particular area. In emergency blood needs, patients can rapidly search for blood banks or hospitals that match a specific or related blood group and contact them via the app. Dutta et al. (2018) designed an integrated Blood Donation Camp Management system to manage end to end blood management process. In this paper, a particular emphasis is placed on the localized blood donation campaign management, in which campaign activities are arranged according to the donor's preferences for time, place, date, etc. The number of donors expected for a scheduled campaign can be estimated using donor registration data.

MATERIALS AND METHODS

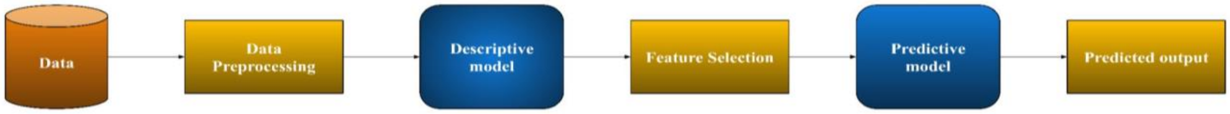


Figure 1. The proposed method

This study has followed the data mining standard approach. The overall process is shown in figure 1. This section describes them individually.

Data Collection

Blood donors' data was legally collected from a small voluntary organization named “Maizbhandari Shah Emdadia Blood Donors’ Group”. The dataset included 8 independent variables such as ID, people of different ages, blood types, genders, last donation date, first donation date, donation count, and city in Bangladesh. By analyzing all independent variables, the dataset includes one dependent variable class to find the most probable donor. If a person will donate blood, then he/she will be classified as 1 otherwise 0. Table 1 shows the dataset attribute descriptions.

Maizbhandari Shah Emdadia Blood Donors’ Group consists of 700 voluntary blood donors. Members of this group are typically residents of Bangladesh's Chattogram city and surrounding areas. The organization has 700 donors' phone numbers and addresses in its database. We contacted 100 donors for this study through physically, phone calls, or social media and gathered data based on the independent variables listed in Table 1. So, the sample size is 100 people, while the population is 700.

Table 1. Attributes of the collected dataset

Variable	Type	Independent/ Dependent	Description
ID	Numeric	Independent	Donor ID
Blood type	Categorical	Independent	The blood type: A+, A-, B+, B-, AB+, AB-, O+, O-
Gender	Categorical	Independent	Male(M), Female(F)
Age	Numeric	Independent	Donor age

Donation count	Numeric	Independent	Number of donations
First donation date	Date	Independent	Date of the first donation
Last donation date	Date	Independent	Date of the last donation
City	Categorical	Independent	Donor present address
Class	Binary	Dependent	1, if a person will donate blood else 0

Data Preprocessing

Feature engineering and, data normalization techniques are used in data preprocessing. Feature engineering takes the original data and adds new attributes that can help modeling. In our dataset, we included two new attributes: 'Difference between first and last blood donation dates' and 'Days since the last donation.' The city variable has been omitted because all of the samples were taken from Chattogram inhabitants. We obtained the following variables as a result of feature engineering:

ID, Blood type, Gender, Age, Donation count, Class, and

- Difference between first and last blood donation dates: The time interval between a donor's first and last blood donation.
- Days since the last donation: The time interval between the current date and the donor's last blood donation. The current date refers to the period during which the research is ongoing.

Descriptive Model

The descriptive model investigates the behavior of donors by analyzing data patterns. We used descriptive statistics and the two-step clustering method to develop the descriptive model. The two-step clustering approach is divided into two steps, the first of which is the pre-clustering of all records. The pre-clusters are sequentially combined in the second stage using a hierarchical agglomerative cluster technique. This approach automatically selects the number of clusters. The whole process was analyzed via software IBM SPSS Statistics for Windows, version 25.0.

Predictive Model

The predictive model predicts whether a donor will donate blood or not during an emergency time by analyzing their past data. This study analyzed the independent variables of previously collected data and divided the blood donors into two categories, most likely (1) and unlikely (0). After that, a training set (80%) and a testing set (the remaining 20%) were created from the gathered data. The data patterns were taught to the model using the training set. The testing set was used for model evaluation.

Logistic regression, J48 decision tree, and Random Forest data mining algorithms are utilized to construct three predictive models. The best prediction model is then determined by comparing these models. All of these methods have been carried out with the tool Weka, version 3.8 (Hall, 2009).

RESULTS

This section illustrates the outcomes of the two models that were developed in this research.

Findings of the Descriptive Model

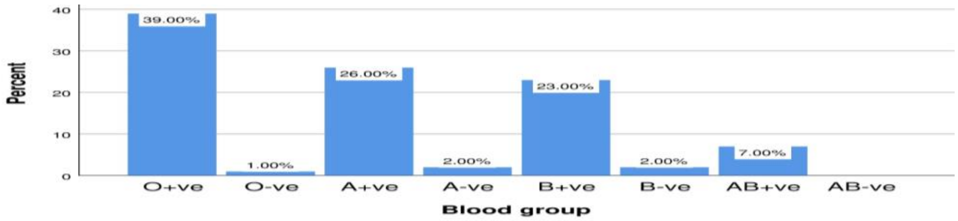


Figure 2. The percentage of donors’ blood group

Descriptive statistics analysis: The highest percentage of blood groups in our samples is O +ve, as shown in figure 2. All the positive blood groups are most common. Negative blood groups are rare, especially AB-ve. That blood group donor was not found in our samples.

The age of blood donors ranges from 18 to 56 years old in our dataset. Donors are an average of 29.22 years old. This explores the fact that the vast majority of donors in our samples are young, making them the most probable donors. The minimum and maximum donation amounts are 0 and 50, respectively. 6.79 is the average number of donations. This shows that the majority of donors in our samples are repetitive, implying that they are the most probable donors.

Figure 3 shows the donation intervals of blood donors. The left one in figure 3 shows the difference in years between the last and first donation, with the x-axis scaled to years from days. The majority of donors, as seen on the left histogram, are between the ages of 0 and 5. 4.2 years is the average. This proves that our dataset has experienced donors who have been donating blood for an average of 4.2 years.

The difference between the current date and the last donation is shown in figure 3's right histogram. The average is 1.2 years in this scenario. This indicates that the vast majority of the donors have given blood recently, i.e. they are the most recent donors.

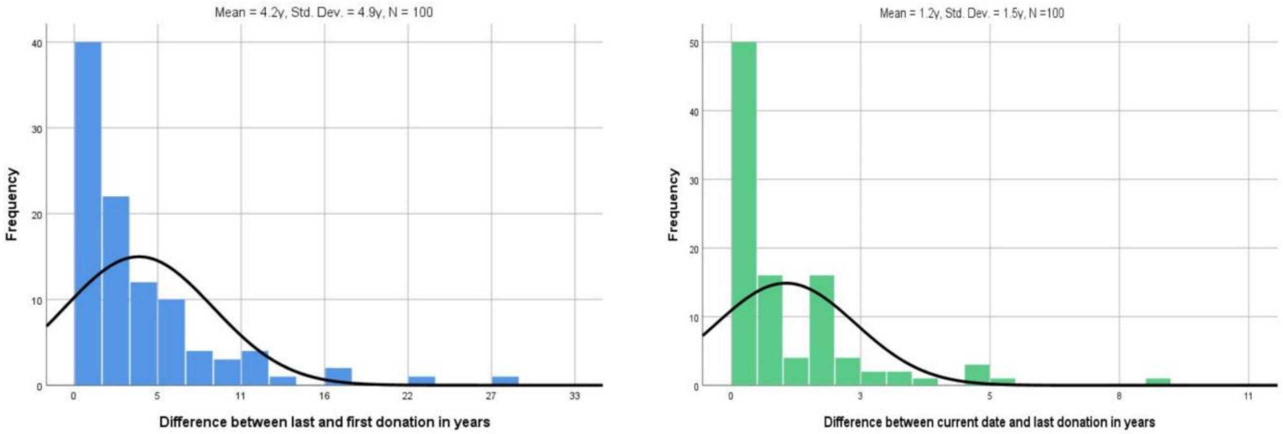
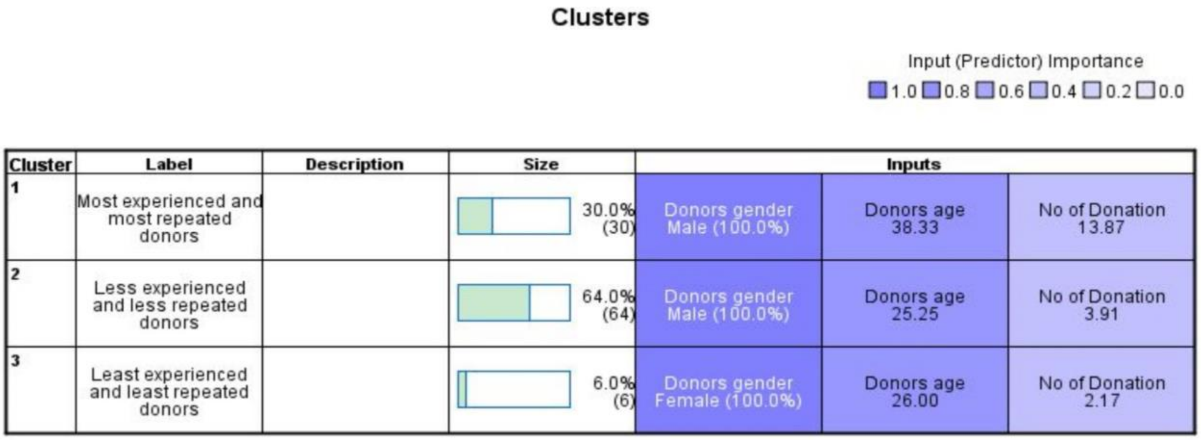


Figure 3. Blood donors’ time interval of donation: (a) The time interval between a donor's first and last blood donation (left). (b) The time interval between the current date and the donor's last blood donation (right).

Cluster analysis: As previously stated, we employed a two-step clustering algorithm to group the samples in order to get diverse behaviour donors based on the three independent variables: donor's age, gender, and number of donations. The results of the clustering algorithm reveal that the study dataset is segmented into three clusters:

- Cluster 1 has 30 samples of blood donors with an average age of 38.33 years, an average blood donation of 13.87, and all of them are men. This suggests that donors in cluster 1 have the most experience and are the most repeated.
- Cluster 2 has 64 samples of blood donors with an average age of 25.25 years, an average blood donation of 3.91, and all of them are men. This investigates the fact that donors in cluster 2 have less experience and are less repeated than cluster 1.
- Cluster 3 contains six blood donors, all of them are women and have an average age of 26 years and a blood donation of 2.17. The fact that donors in cluster 3 have the least experience and are the least repeated than clusters 1 and 2 is investigated.

After analyzing three clusters, it becomes evident that men donate significantly more blood than women. Hence, men are most probable to blood donation. Figure 4 depicts cluster results, with clusters 1, 2, and 3 labeled as most experienced and most repeated, less experienced and less repeated, and least experienced and least repeated donors, respectively.



Clusters

Input (Predictor) Importance

▨ 1.0 ▨ 0.8 ▨ 0.6 ▨ 0.4 ▨ 0.2 □ 0.0

Cluster	Label	Description	Size	Inputs		
1	Most experienced and most repeated donors		30.0% (30)	Donors gender Male (100.0%)	Donors age 38.33	No of Donation 13.87
2	Less experienced and less repeated donors		64.0% (64)	Donors gender Male (100.0%)	Donors age 25.25	No of Donation 3.91
3	Least experienced and least repeated donors		6.0% (6)	Donors gender Female (100.0%)	Donors age 26.00	No of Donation 2.17

Figure 4. Cluster analysis of blood donors

Findings of the Predictive Model

After evaluating the descriptive model's findings, we classified the samples into donors with the highest probability of blood donation and donors with no possibility of blood donation using the following rules:

- If the number of donations of any donor is less than two then he or she was not recognized as a possible blood donor.
- If the value of the variable 'Days since last donation' is fewer than 56 days, he or she was not identified as a possible blood donor. Since a donor has to wait at least 56 days for donations of whole blood ("Eligibility Requirements," n.d.).

- If a donor's time since their last donation exceeds five years, he or she was not identified as a possible blood donor.

Table 2. Evaluation results of predictive models

Performance metric	Random Forest	J48 Decision Tree	Logistic Regression
Precision	100%	96%	85%
Recall	100%	95%	85%
F1 Score	100%	95%	85%
Accuracy	100%	95%	85%
MCC	100%	89%	63%

Apart from these three criteria, the remainder were chosen as the most probable blood donors. According to these rules, we found that out of 100 samples there were 73 individuals that were most likely to donate blood and that there were 27 who were unlikely. These samples were then distributed as 80% train set and 20% test set. With this train data, three models were trained with three data mining algorithms to categorize the donor by looking at its characteristics. After training, the performance of the models is checked with 20% unseen data or test data. It was checked whether the models will be able to accurately classify new blood donors into the most likely or unlikely donors. Table 2 shows the results.
The prediction performance of the models was examined through precision, recall, f1 score, accuracy, and Matthews Correlation Coefficient (MCC) metrics. In table 2 we can see that the Random Forest model gives us 100% accurate results, i.e. it can predict 100% accurately whether a donor will donate blood or not. The reason for giving 100% correct results is that our sample size is small. If the sample size is larger, the accuracy will be further reduced. Since the Random Forest model achieves 100% accuracy across all four metrics, we choose Random Forest to be our preferred predictive model from the above three models.

DISCUSSION

When a patient needs emergency blood, finding the right blood donor in a short amount of time might be difficult. Using data mining approaches, this research article demonstrates how to locate the most potential blood donors from a group in a short amount of time. In this study, we found that men donate significantly more blood than women. Blood donation rates are substantially greater among experienced blood donors or those who have donated blood for a long time. So, we have developed a predictive model by identifying experienced and repeated blood donors. This model can automatically predict the most potential blood donors from a blood donor group in less time which would take more time to do manually. This helps to minimize the time required to find the right donor. By contacting these potential blood donors, the patient can get his/her desired blood in a shorter period of time, and thus the time gap between blood demand and supply is also minimized.

However, we have faced several limitations while doing this research. The biggest problem is the lack of data. We couldn't reach the number of donors that we intended to contact. The major obstacle to data collection is time and budget. Another problem is that during the Covid-19 pandemic the possible blood donors may not be available at the emergency time due to lockdown. This case can also happen in terms of any natural calamities.

The following actions can be recommended to improve Bangladesh's or other developing countries' present blood donation scenario:

- Encourage individuals to donate blood, particularly women, through social awareness programs, seminars, blood donation campaigns, and social media. Every year, the Maizbhandari Shah Emdadia blood donors' group organizes a number of blood donation campaigns and social awareness activities to achieve this goal.
- Collect and maintain a large database of blood donors and seekers.
- Create a web-based blood donation application that incorporates data mining, machine learning, and artificial intelligence technology for effective data management.

REFERENCES

Adetayo Olaniyi, A., & Olufunto Adedotun, K. (2018). Artificial Intelligence in Aircraft Docking: The Fear of Reducing Ground Marshalling Jobs to Robots and Way-Out. *American International Journal of Multidisciplinary Scientific Research*, *1*(2), 25–32. https://doi.org/10.46281/aijmsr.v1i2.185

Agarwal, K., Gupta, M., Gupta, K., Khan, A., & Nallakaruppan, M. K. (2019, March). Blood Transfusion System Using Data Mining Techniques and GRA. *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, 1143–1147. https://doi.org/10.1109/icaccs.2019.8728401

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. *International Journal of Academic Engineering Research (IJAER)*, 3 (10), 1-7.

Alkahtani, S. A., & Jilani, M. (2019). Predicting Return Donor and Analyzing Blood Donation Time Series using Data Mining Techniques. *International Journal of Advanced Computer Science and Applications*, 10(8).

Al-Masum Molla, M. (2017, June 14). Number of blood donors on rise. Retrieved June 10, 2021, from https://www.thedailystar.net/backpage/number-blood-donors-rise-1419814

Ashoori, M., & Taheri, Z. (2013, August). Using clustering methods for identifying blood donors behavior. In *5th Iranian Conference on Electrical and Electronics Engineering (ICEEE2013)* (pp. 4055-4057).

Bangladesh is still to meet the demand of safe blood supply. (2017, June 14). Retrieved June 11, 2021, from https://www.who.int/bangladesh/news/detail/14-06-2017-bangladesh-is-still-to-meet-the-demand-of-safe-blood-supply

Bhardwaj, A., Sharma, A., & Shrivastava, V. K. (2012). Data mining techniques and their implementation in blood bank sector–a review. *International Journal of Engineering Research and Applications (IJERA)*, *2*(4), 1303-1309.

Blood safety and availability. (2020, June 10). Retrieved June 11, 2021, from https://www.who.int/en/news-room/fact-sheets/detail/blood-safety-and-availability

Dutta, L., Maji, G., Ghosh, P., & Sen, S. (2018, October). An Integrated Blood Donation Campaign Management System. *Advances in Intelligent Systems and Computing*, 133–143. https://doi.org/10.1007/978-981-13-1540-4_14

Eligibility Requirements. (n.d.). Retrieved July 7, 2021, from https://www.redcrossblood.org/donate-blood/how-to-donate/eligibility-requirements.html

Fahad, A. A. A. (2019). Design and implementation of blood bank system using web services in cloud environment. *International Journal of MC Square Scientific Research*, *11*(3), 09-16.

Gulinac, M., Dikov, D., & Velikova, T. (2020). Epidemiological and Morphological Characteristics of Urothelial Bladder Cancer in a Bulgarian and a French Sample of Patients. *American International Journal of Multidisciplinary Scientific Research*, *6*(1), 1–5. https://doi.org/10.46281/aijmsr.v6i1.547

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software. *ACM SIGKDD Explorations Newsletter*, *11*(1), 10–18. https://doi.org/10.1145/1656274.1656278

Hashim, S. A., Al-Madani, A. M., Al-Amri, S. M., Al-Ghamdi, A. M., & Nahla, B. S. B. (2014). Online Blood Donation Reservation and Management system In Jeddah. *Life Science Journal*, *11*(8).

KIRCI, P., AKTAS, S., & SEVINC, B. (2020, June). Analyzing Blood Donation probabilities and number of possible donors. *2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 1–4. https://doi.org/10.1109/hora49412.2020.9152872

Lakshmi, B. V., & Kumar, B. R. (2018). Leveraging Technologies to Redefine Business: Technology Perspective. *American International Journal of Multidisciplinary Scientific Research*, *1*(3), 1–3. https://doi.org/10.46281/aijmsr.v1i3.187

Sahariar Hasan Jiisun, M., Akter Rupa, R., Hussain Chowdhury, M., Mushrofa, H., & Hoque, M. R. (2019). Blood Donation Systems in Bangladesh: Problems and Remedy. *International Journal of Business and Management*, *14*(8), 145. https://doi.org/10.5539/ijbm.v14n8p145

Serteva, D., Poryazova, E., & Velikova, T. (2019). Endometriosis Locations and Coexistence with other Uterine Conditions in a Bulgarian Sample of Patients. *American International Journal of Multidisciplinary Scientific Research*, *5*(2), 5–9. https://doi.org/10.46281/aijmsr.v5i2.255

Tithila, K. K. (2019, June 14). Donate blood, save life. Retrieved June 18, 2021, from https://www.dhakatribune.com/bangladesh/event/2019/06/15/donate-blood-save-life

Velikova, T., Velikov, T., & Mihailov, G. (2018). Higher Serum Levels of Stromelisyn 2 (MMP10) but not Matrilysin (MMP7) in Patients with End-Stage Chronic Kidney Disease on Chroniodialysis. *American International Journal of Multidisciplinary Scientific Research*, *4*(1), 17–21. https://doi.org/10.46281/aijmsr.v4i1.196

World Health Organization. (2010). Towards 100% voluntary blood donation: a global framework for action.

Yeh, I. C., Yang, K. J., & Ting, T. M. (2009). Knowledge discovery on RFM model using Bernoulli sequence. *Expert Systems with Applications*, *36*(3), 5866–5871. https://doi.org/10.1016/j.eswa.2008.07.018

Copyrights