International Journal of Applied Engineering Research
ISSN 0973-4562 Volume 10, Number 7 (2015) pp. 18503-18514

http://www.ripublication.com

A Study on Automation of Blood Donor Classification and Notification Techniques

Arunkumar Chinnaswamy, Gurusankar Gopalakrishnan, Kamal Kiran Pandala Venkata, Shabala Natarajan

Department of Computer Science
Amrita School of Engineering, Amrita Vishwa Vidyapeetham
Amrita Nagar, Coimbatore, Tamil Nadu 641112 India
(arunkumar.chinnaswamy, gurushankar.gopalki, kamalkiran1993, shabalanatarajan)@gmail.com)

1. ABSTRACT

The increasing demand for sophisticated, intelligent systems in the field of healthcare leads to a need for introduction of automation of processes. The area of transfusion medicine, specifically blood donation services require this implementation at the earliest. The present situation is one where most processes in blood donation services are manual and the demand for blood is constantly on the rise, augmented by declining donation rates. Hence, an intelligent system that can integrate major operations involved, make efficient decisions, improve communication and introduce significant agility is highly crucial. A system of this sort would involve a uniform labelling standard for blood packet identification, machine learning algorithms for efficient donor selection and a notification system. A general overview and comparison between various such techniques in existence, independently or in combination, is presented in this review.

Keywords: blood, classification, donation, intelligent systems, notification

2. INTRODUCTION

The requirement for blood is steadily increasing as a result of increasing population and also, advancement in clinical medicine. On the contrary, the number of voluntary donors is decreasing over the last few years. Recent findings bring to light the fact that a majority of the new donors fail to donate a second time. It is not economically viable to continually check for donors who fail to return for a subsequent donation. Retention of donors who pass the initial health checks is vital to the strategy of increasing blood donation. However, blood donation services can alter their strategies to rope in donors only if they can understand the factors that influence their donation behaviour. Finding and understanding these factors would help the blood donation services to come up with newer and more efficient strategies that would increase the return rate of donors [1]. An analysis conducted in [2] shows that the factors influencing the decision to donate blood are complex.

This creates an impending need for an intelligent automation mechanism that can, with minimalistic human intervention, monitor blood stock, select potential blood donors and gather them based on blood requirement. The incorporation of information and computer technology can help in decreasing the workload of blood centres and in the analysis of the contributing factors. In this paper, we review and contrast various existing implementations and previous research done on this aspect. Further, we identify the root components of the proposed intelligent system and offer a study on the potential methods of its implementation.

3. RELATED WORK

Significant research and work has been previously done to address and the problem of optimal blood distribution and enhancement of the efficiency of blood donation process. Web-based systems are the most common with significant work on automated online blood bank databases and real time monitoring systems [3][4]. While notification systems involving various communication technologies have been studied, research heavily supports mobile communication as the most viable method. Further, messaging services (SMS) and smartphone applications are the most popular [5][6][7][8].

Machine Learning techniques form the backbone of the proposed mechanism. Studies have been conducted on various implementation techniques like Decision Trees, Random Forests, Bayesian Networks, Neural Networks, Extreme Learning Machines, clustering techniques and their accuracies [9][10][11]. Further, there has been research and proposal of analytical models and systems to enhance blood donor selection and blood collection [12][13]. Donor selection has been done by taking into account altruism as the most contributing factor [14].

4. MACHINE LEARNING TECHNIQUES

Machine learning algorithms are widely used to render potential solutions to real-life problems in the field of computer science. Many algorithms have been implemented and compared for performance and efficiency by employing them on various datasets, most of them belonging to medical fields. Decision Trees (DT), Random Forests (RF), Artificial Neural Networks (ANN), Bayesian Networks, Support Vector Machines (SVM) and Gaussian Processes (GP) are among the most frequently used. A comparative study on several of these methods indicates what algorithms suit a given data set the best. A dataset of ICU patients with head injuries, classified using DT reveals that the implementation is not only easily comprehendible but also has a moderately high resistance to noise and errors. An ensemble of DT is RF, which shows better performances and hence is more suited for datasets where scalability comes into play [15]. Neural Networks are found to be robust to errors and hence are suited for noisy data such as live inputs. Their lack of interpretability favours the implementation of DT over ANN.

The process of record keeping in hospitals, which is now mostly manual and error-prone, becomes simpler, more efficient and with a reduced monitor-replenish cycle by a Markov chain implementation [12]. A simple M/M/I/k queueing model is utilized to study the suitable rate of acquisition as a function of consumption, minimum amount which is required and the maximum available storage capacity. This model applies for a single component of blood, for each component a separate model must be devised. Different thresholds and activation of notification of system based on its values.

M M Mostafa conducts a research analysis and tries to profile blood donors for optimized recruitment [14]. Blood banks and transfusion centres depend entirely on volunteer blood donors but it is shown that it is a challenge to find considerable number of volunteers. Research conducted attempts to identify factors that discriminate donors from non-donors. Conclusion is drawn from various psychological, conceptual, behavioural and related literature, hypotheses supporting correlations between said factors and blood donation tendencies.

Multiple Machine Learning approaches are used to profile donors. Comparisons between Linear Discriminant Analysis (LDA) and Multi-layer Perceptron (MLP) implementations and various other Machine Learning approaches used to profile donors are done, of which the artificial neural network approach of PNN (probabilistic neural networks) shows the highest accuracy of 100% [14]. Depending on the algorithm implemented, the accuracy with which the classification of potential donors changes and using the results of the classification, recruitment of donors to satisfy the blood requirement is possible.

A Adewumi et al, presents an optimization problem, modelling the blood assignment in scenarios of distribution ,which aims at minimizing the quantity of blood that is to be imported from other sources, hence reducing blood shortage by effective management of given resources [16]. The general model of a blood bank is assumed to be similar to one shown in Fig 1. (Fig1)

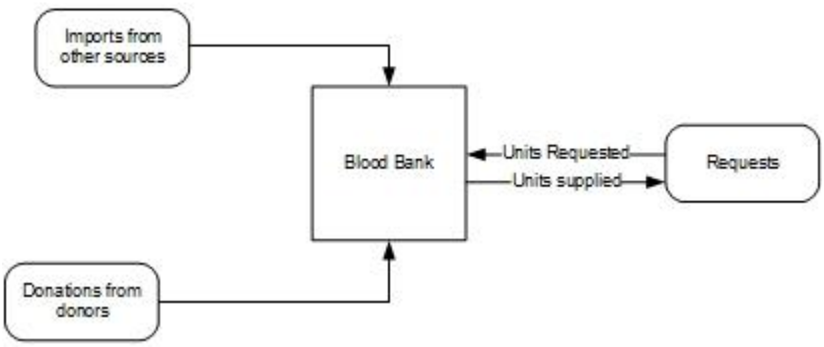

Figure 1 General Model of a Blood Bank.

The proposed optimization solution is a modification of the model in Fig 1 as shown in Fig 2. (Fig2)

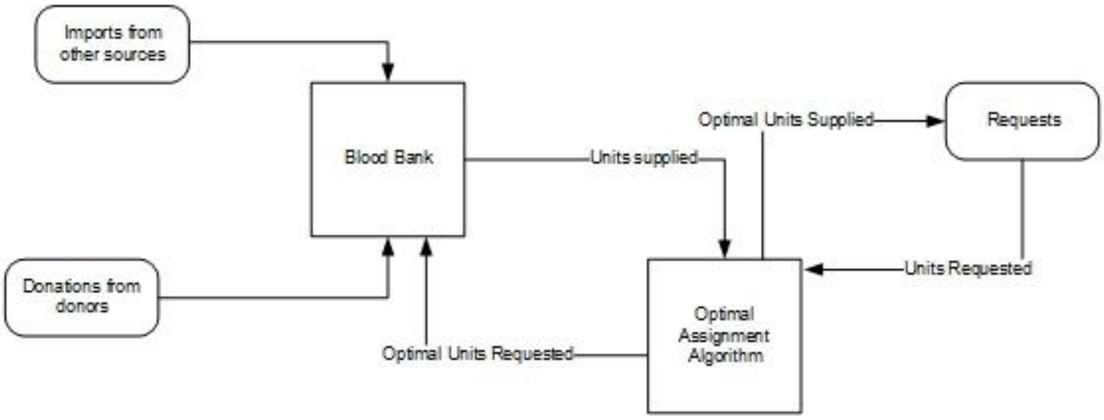

Figure 2 Proposed Optimization Solution.

It suggests a new mathematical model that is not as restrictive as the ones proposed in [17], referred to as the Simple Linear assignment. The simple linear assignment algorithm schedules requests immediately, does not allow cross match for blood types and is used as the benchmark.

The optimization problem is modelled into a multiple knapsack solution where it is experimented to see whether cross matching compatible blood types in order to satisfy blood requests can be used to stabilize the proportion of blood type usage within the blood bank as well as minimize scenarios where blood has to be imported. The model is elaborated with constraints on blood compatibility and blood types assigned values based on how many blood types it is compatible with.

Results and conclusions in [16] show that the multiple knapsack solution fares better than the simple assignment solution in reducing the number of units that is to be imported. There were a few negative side effects where the MKA used more of the most valuable blood type (O-) due to it getting substituted in cross matching of compatible blood types. This efficient model for managing and distribution of blood utilizes various optimization strategies to show that optimization is possible with cross matching.

Risk assessment in medical datasets demands two factors- the probability of an event occurring and a measure of impact severity by the occurrence of the event. This is satisfied with a BN implementation. A case study on risk calculation from real time data collected from End Stage Renal Disease patients shows that different sources of data and knowledge can be integrated using BN [15]. This is of significance in cases where dependencies of variables is crucial for decision making, as is in the case of risk assessment. Further, loss functions are important for risk estimations and classifications involving real-time data. These can be easily included in a BN implementation. BN allows integration of risk estimation with decision making; making it a favourable choice for most real- time implementations.

SVMs have also been applied for classification of datasets from medical domain. In the prediction of tacrolimus concentration in blood in liver transplant patients from an ICU dataset, SVMs require less inputs and outperformed multivariate linear regression [18]. They also show a better generalization behaviour than logistic regression in the prediction of the depth of infiltration in endometrial carcinoma [19]. A study of acute leukaemia expression profiles using Support Vector Regression reveals that the interclass boundary is very specific to those in the training dataset. Thus, an SVM will not generalize well when presented with new gene profiles and hence will cause significant over-fitting. The study on gene expression measurements shows that algebraic combination of multiple kernels allows a single SVM to perform classification on multiple data types simultaneously [20].

Most instances of data in real-life has a high degree of uncertainty attached to it. This creates complexity in their study and analysis; which hence, dictates the need for a probability density function. Extreme Learning Machine (ELM) algorithms are ones that aid in binary and multi-class classification of uncertain data. They reduce the time required to train neural networks. ELMs are better suited for online and real time applications because of their ability to automatically detect network parameters.

Experiments on diabetes datasets in [21] reveal that a Phase Encoded ELM implementation (PE) has a higher classification accuracy than SVM or LS-SVM. Further, training time comparisons over binary class and multiclass classification show the value for LS-SVM to be almost tenfold as the ones for ELM. Hence, it can be concluded that ELM implementations significantly reduce training time for binary as well as multiclass classification. A study of an artificial case of approximation of sinC function with noise suggests that the response to an unknown external stimuli would be faster by a SLFN (ELM) than SVM [22].

A system is proposed in [13]that makes use of clustering and classification algorithms to determine the discrepancies in blood donation behaviour among the existing donors and predicts their intentions towards blood donation by understanding the contributing factors. These factors are then utilized to design a strategy that would lead to the increase in voluntary blood donation frequency. The proposed system has the flow depicted in Fig.3. (Fig3)

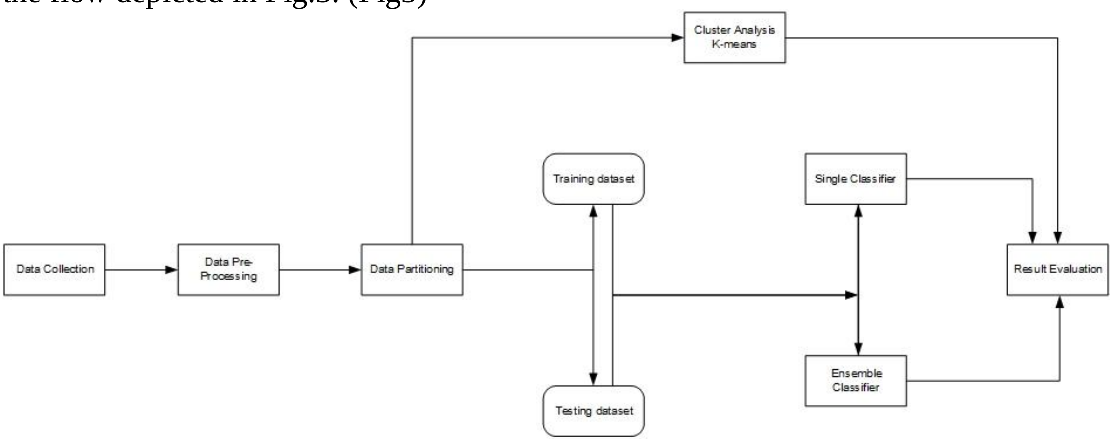

Figure 3 Proposed System Flow.

The system adopts a donor dataset from a blood transfusion centre in Hsin-Chu City, Taiwan [23]. First, donors are clustered based on their opinion on donation using lazy learners. The attributes that influence blood donation are standardized and the tuples are ranked based on the created standard. This facilitates easier interpretation of the results. Hopkins Statistic is used to determine the quality of the resultant clusters [24]. ANOVA is used to test if there exists any significant difference among the attributes. But the clusters formed still do not explain the behaviour.

Supervised data mining tools can be used to understand the nature of the clusters formed by the unsupervised clustering algorithms. The attributes that are best able to discriminate are selected and used to construct a classification algorithm. The techniques Decision Tree (DT), Naïve Bayes (NB), NBTree and their ensembles are used. The data is split into subsets according to the dataset partitioning step of the system. The training data is fed into the system which builds a preliminary classification model that uses the testing data to test accuracy. By better understanding the factors that influence a donor’s intentions towards blood donation, the blood donation services can implement various strategies that would facilitate efficient donor selection and collection of blood.

5. NOTIFICATION MECHANISMS

The merging of latest, widely used trends of communication in the field of healthcare plays a role of significance in the form of notification systems in scenarios that require emergency communication. Web-based and mobile communication systems are amongst the most popular modes adopted. India has 11 ongoing projects with respect to mobile communications for eHealth solutions [25].

A common and easy way to implement a notification system is via the use of the Short Messaging Service (SMS). SMS, initially designed as a means of inter-person communication, has been extended to be utilized as an alerting system. SMS based alert systems are used across domains varying from tissue culture monitoring in biology [26] to flight delay alert system in business. A case study on automation of a stationery request procedure [7] shows that SMS notification is simple in its implementation and influences the customer feedback positively. Moreover, it reduces the life cycle of a request drastically, making the implementation efficient.

Smart Blood Query (SBQ) [6] focuses on keeping the information secure, the donors and recipients anonymous and follows the guidelines provided by WHO for safe blood management. The proposed system is developed using RapidSMS, an open source SMS-based tool [27]. Notification is automated when blood stock falls below a threshold value. Periodical polling is required to check the blood stocks and calculate the threshold trigger. This requires establishing a set of parameters for each blood type:

- Saturation Point: The volume with which a hospital can operate ideally without any requirement to replenish their stocks. A volume above this point contributes to wastage of blood due to expiry.
- Sufficiency Point: The minimal volume of blood stock with which a hospital can operate. Any value less than this calls for emergency re-stocking from

other hospitals or blood banks.

- Filling Point: The maximum volume of blood that can be stocked at a given time. A stock of this volume has a high probability of contributing to wastage of blood due to expiry as all of it may not be used.
- Null Point: A hypothetical situation where there is absolutely no stock of a given blood type.

The first two values are dynamic. They are estimated from the past blood stock usage data and vary with time and demand. The median of the blood stock data is also calculated, which lies in between the saturation (upper) level and the sufficiency (lower) level. The following describes the operations of the system at various possible ranges of the blood stock:

- Saturation Point < Present Volume < Filling Point: No blood collection required.
- Median < Present Volume < Saturation Point: Notification sent to a group of most eligible and optimal donors (those with a relatively high probability of turning up for donation) to collect blood.
- Sufficiency Point < Present Volume < Median: Notification sent to a larger group of eligible and optimal donors to quickly increase the blood stock level above the median value.
- Null Point < Present Volume < Sufficiency Point: This level is never attained because of constant polling and updating of blood stock.

The number of donors to be potentially notified depends upon a sigmoid function. Lesser the deficit in blood stock, a fewer donors are notified, but with increasing deficit, the number of donors notified increases exponentially. Arrest of the notification mechanism takes place once the value of stock goes above the saturation point, thereby minimizing wastage due to expiry. Fig. 4 depicts different thresholds. (Fig4)

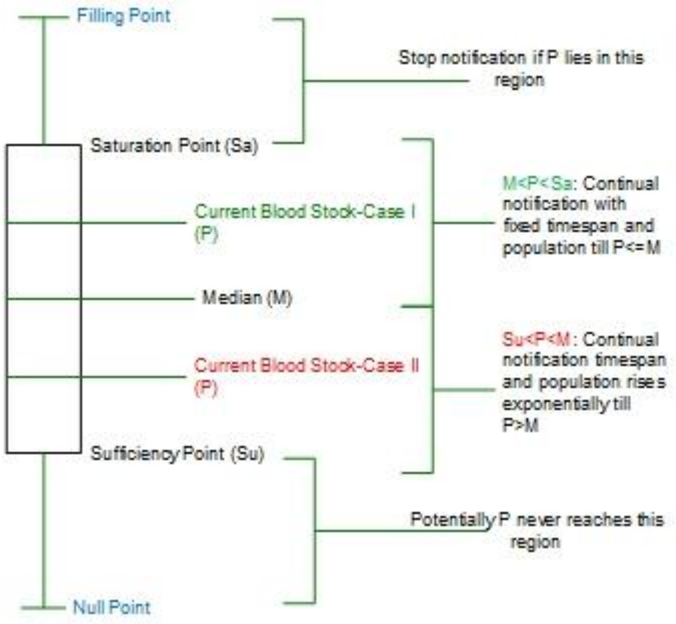

Figure 4 Threshold Level Calculation.

Ranking of donors based on availability and location increases the probability of a positive response. This is complemented by bulk SMS dispatch in case of emergency. This two tier notification mechanism suggests a definite way of boosting response probability.

Virtual Blood Bank Project (VBB) [28] suggests that even though web based blood donor matching facilities for recipients are plenty, they do not work in the cases of emergencies that might arise and that they are entirely dependent on the Internet, which is a facility which is not as ubiquitous as cell phones in emerging economies like India, drawn as a conclusion from [29].

The Automated Online Blood Bank Database (AOBBD) discussed in [3], can primarily be classified as a web based blood distribution solution and does not take into account this lack of Internet facilities, but circumvents around the issue by implementing a direct contact feature, wherein the donor and the recipient can connect with each other, with a phone call, without the need for an Internet connection.

It is pointed out that a basic online blood bank database would be of little help without a call routing facility, which is proposed using asterisk hardware [3]. It is also claimed that a system that provides direct contact via call helps with a more immediate response than those employing SMS based or just internet based databases. In general, we find major similarities in the operation of the VBB project [29] and AOBBD [3] .The generalized model is shown in Fig 5. (Fig5) The major differences are in the implementation of the architecture, request communication to the server, and connection establishment between recipients and donors.

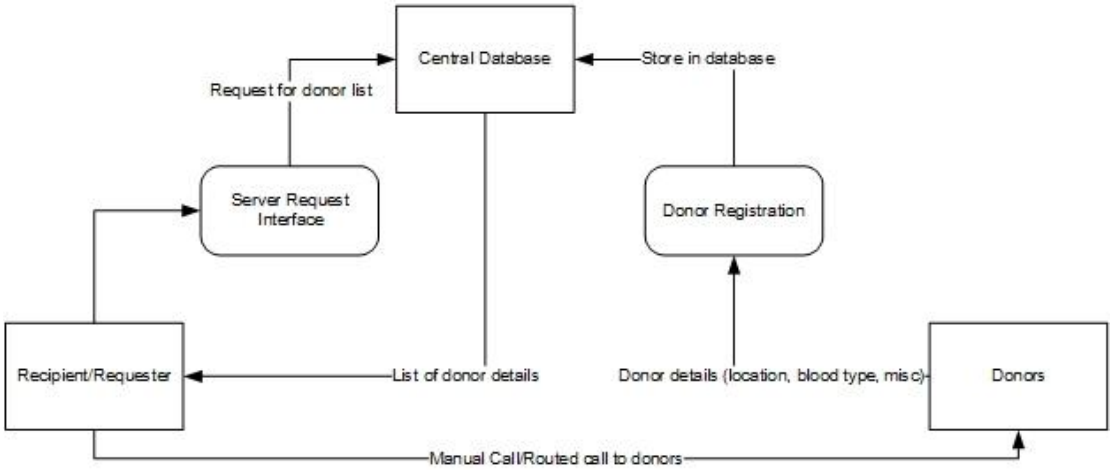

Figure 5 Generalized model of the system.

mHealth, a blood donation service enhanced by SMS has been implemented in Bangladesh and provides great insights at how mobile communication and SMS can be used as the notification system for a blood donation service [8]. The study considers cellular and internet infrastructure in Bangladesh and concludes that an internet based system is not the optimum solution for a blood donation service, given the ICT context of the nation. This case can be extrapolated to other South and South-East Asian nations, given the exponential rise in the number of mobile users. Thus, given the feasibility of deployment, wide span of reach, and economic factors, a mobile SMS based notification system would be the most optimal solution. Further, such a system has very low requirements for its functioning. A GSM modem and SIM card suffices for the system to be deployed.

It also identifies the factors affecting voluntary donation through surveys and discussions. Performance of the proposed system has been assessed by gender-based ranking. Aside from inherent pros and cons in systems discussed above, we find that all the systems aim to increase blood availability and reduce the latency in response. The huge support for the proposed idea suggests the need for such notification systems in developing and enhancing services in healthcare.

6. CONCLUSION

We present a survey of papers relating to automation in the area of transfusion medicine for aiding and enhancing donation and distribution. Machine Learning techniques are the most viable candidates for implementing ranking and selection of donors, given their efficiency, accuracy and ease of implementation. The development of a stable, error free communication mechanism is very vital in the system proposed. SMS is the favoured option in this case, given the widespread use of cellular phones and hence ease of access. While web-based systems are quite efficient, they prove to have a lesser reach than SMS, given the context that has a high degree of emergency and hence gives importance to time taken to reach out to donors.

The basic components of the system proposed in this paper are thus, an intelligent Machine Learning algorithm with a fair accuracy and a sophisticated notification system. By the integration of the above components, we can achieve a machine that would aptly satisfy the needs of blood donation services.

7. REFERENCES

[1] Ringwald, J., Lange, N., Rabe, C., Zimmermann, R., Strasser, E., Hendelmeier, M., Strobel, J., and Eckstein, R., 2007, "Why do some apheresis donors donate blood just once?", Vox Sang, 0(0), pp. 070824040958001-???.

[2] France, J., France, C., and Himawan, L., 2007, "A path analysis of intention to redonate among experienced blood donors: an extension of the theory of planned behavior", Transfusion, 47(6), pp. 1006-1013.

[3] Arif, M., Sreevas, S., Nafseer, K., and Rahul, R., 2012, "Automated online Blood bank database", 2012 Annual IEEE India Conference (INDICON).

[4] Guangpeng, L., Zhongwen, G., Song, X., and Wenli, P., 2009, "Web-based real-time monitoring system on cold chain of blood", 2009 IEEE Intrumentation and Measurement Technology Conference.

[5] Singh, R., Bhargava, P., and Kain, S., 2007, "Smart Phones to the Rescue: The Virtual Blood Bank Project", IEEE Pervasive Comput., 6(4), pp. 86-89.

[6] Rahman, M., Akter, K., Hossain, S., Basak, A., and Ahmed, S., 2011, "Smart Blood Query: A Novel Mobile Phone Based Privacy-Aware Blood Donor Recruitment and Management System for Developing Regions", 2011 IEEE Workshops of International Conference on Advanced Information Networking and Applications.

[7] Samsudin, N., Khalid, S. A., Yusoff, A. M., Ihkasan, M., and Senin, Z., 2011, "Procedure automation with immediate user notification: A case study", 2011 IEEE Symposium on Business, Engineering and Industrial Applications (ISBEIA).

[8] Islam, A., Ahmed, N., Hasan, K., and Jubayer, M., 2013, "mHealth: Blood donation service in Bangladesh", 2013 International Conference on Informatics, Electronics and Vision (ICIEV).

[9] Xue, M., and Zhu, C., 2009, "A Study and Application on Machine Learning of Artificial Intellligence", 2009 International Joint Conference on Artificial Intelligence.

[10] Meyfroidt, G., Güiza, F., Ramon, J., and Bruynooghe, M., 2009, "Machine learning techniques to examine large patient databases", Best Practice & Research Clinical Anaesthesiology, 23(1), pp. 127-143.

[11] Sun, Y., Yuan, Y., and Wang, G., 2014, "Extreme learning machine for classification over uncertain data", Neurocomputing, 128, pp. 500-506.
[12] Boppana, R., and Chalasani, S., 2007, "Analytical Models to Determine Desirable Blood Acquisition Rates", 2007 IEEE International Conference on System of Systems Engineering.
[13] Lee, W., and Cheng, B., 2011, "An Intelligent System for Improving Performance of Blood Donation", Journal of Quality, 18(2), pp. 173-185.
[14] Mostafa, M., 2009, "Profiling blood donors in Egypt: A neural network analysis", Expert Systems with Applications, 36(3), pp. 5031-5038.
[15] Kenett, R., and Lavi, Y., 2014, "Integrated management principles and their application to healthcare systems", Sinergie rivista di studi e ricerche.
[16] Adewumi, A., Budlender, N., and Olusanya, M., 2012, "Optimizing the assignment of blood in a blood banking system: Some initial results", 2012 IEEE Congress on Evolutionary Computation.
[17] Angelis, V. D., Ricciardi, N., and Storchi, G., 2001, "Optimizing Blood Assignment in a Donation-Transfusion System", Int Trans Operational Res, 8(2), pp. 183-192.
[18] Looy, S. V., Verplancke, T., Benoit, D., Hoste, E., Maele, G. V., Turck, F. D., and Decruyenaere, J., 2007, "A novel approach for prediction of tacrolimus blood concentration in liver transplantation patients in the intensive care unit through support vector regression", Critical Care, 11(4), pp. R83.
[19] Pochet, N., and Suykens, J., 2006, "Support vector machines versus logistic regression: improving prospective performance in clinical decision-making", Ultrasound in Obstetrics and Gynecology, 27(6), pp. 607-608.
[20] Golub, T., 1999, "Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring", Science, 286(5439), pp. 531-537.
[21] Rajesh, R., and Prakash, J., 2011, "Extreme Learning Machines - A Review and State-of-the-art", 2011 International journal of wisdom based computing, 1(1), pp. 35-49.
[22] Huang, G., Zhu, Q., and Siew, C., 2006, "Extreme learning machine: Theory and applications", Neurocomputing, 70(1-3), pp. 489-501.
[23] Yeh, I., Yang, K., and Ting, T., 2009, "Knowledge discovery on RFM model using Bernoulli sequence", Expert Systems with Applications, 36(3), pp. 5866-5871.
[24] Banerjee, A., and Dave, R., 2004, "Validating clusters using the Hopkins statistic", 2004 IEEE International Conference on Fuzzy Systems (IEEE Cat. No.04CH37542).
[25] Vatsalan, D., Arunatileka, S., Chapman, K., Senaviratne, G., Sudahar, S., Wijetileka, D., and Wickramasinghe, Y., 2010, "Mobile Technologies for Enhancing eHealth Solutions in Developing Countries", 2010 Second International Conference on eHealth, Telemedicine, and Social Medicine.
[26] Aziz, N., Muhamad, W., Wahab, N., Alias, A., Hashim, A., and Mustafa, R., 2010, "Real Time Monitoring Critical Parameters in Tissue Culture Growth

Room with SMS Alert System", 2010 International Conference on Intelligent Systems, Modelling and Simulation.
[27] Rapidsms.org, 2015, "RapidSMS: A Free and Open Source SMS Framework" [Online]. Available: https://www.rapidsms.org/. [Accessed: 29- Mar- 2015].
[28] Singh, R., Bhargava, P., and Kain, S., 2007, "Smart Phones to the Rescue: The Virtual Blood Bank Project", IEEE Pervasive Comput., 6(4), pp. 86-89.
[29] Pervasive Computing, April-June 200, IEEE.

8. BIOGRAPHICAL SKETCH

Arunkumar Chinnaswamy did B.E in Computer Science and Engineering at Dr.Mahalingam College of Engineering and Technology, Pollachi and his M.Tech in Computer Science and Engineering at Vellore Institute of Technology, Vellore. He is currently pursuing his Ph.D at Anna university, Coimbatore in the area of biomedical image processing. He is working as an Assistant Professor in the Computer Science and Engineering department of Amrita School of Engineering, Amrita University, Coimbatore since August 2006. Placed in WIPRO technologies, Bangalore, he took up the teaching profession out of his own interest. His research paper in the area of Mobile Communication won the “Best Research Paper award” in the Computer Society of India National Conference in 2009. He is responsible for starting “Institution of Electronics and Telecommunication Engineers”(IETE) student forums in all engineering colleges and polytechnics in Tamilnadu under the jurisdiction of IETE Coimbatore center which expands its wings to more than 12 districts in Tamilnadu. He has also inaugurated student forums and delivered lectures in more than 30 engineering colleges in Tamilnadu

Gurusankar Gopalakrishnan is a final year student of B.Tech Computer Science and Engineering at Amrita School of Engineering, Amrita Vishwa Vidhyapeetham, Coimbatore. His research work includes Graphical analysis for large social networks and Computational creativity in images. His interests lie in Machine learning and Big Data.

Kamal Kiran P V is a final year student of B.Tech Computer Science and Engineering at Amrita School of Engineering, Amrita Vishwa Vidhyapeetham, Coimbatore. He has experience in Web development and his areas of interests include Data Mining and Data Analytics.

Shabala Natarajan is a final year student of B.Tech Computer Science and Engineering at Amrita School of Engineering, Amrita Vishwa Vidhyapeetham, Coimbatore. Her research areas include Statistical Machine learning and Business intelligence. She has experience in working with Web development as well.