# BLOOD DONATION INTERVAL ESTIMATION THROUGH DEEP LEARNING

Parvatibai Genba Moze College of Engineering Wagholi,Pune
Department of Computer Engineering

1. Wruchita Deshmukh
2. Komal Borhade
3. Akbar Shaikh
4. Araju Bansod

Internal Guide:- Prof. Dnyandeo Khemnar

***Abstract* – Blood is a highly valuable commodity that cannot be created in a lab but can only be produced by human body. This is different from other type of drugs and chemicals where most of them can be synthetically produced and provided without the need of any donations. But the complex fluid blood produced by humans cannot be replicated which can be problematic in scenarios where a large amount of blood loss for an individual threatens their survival. For this purpose blood donation has been encouraged vehemently as donating blood can be tantamount to saving a life. But most of the blood donation banks and other institutions prescribe a weight of 6 months between donations which can lead to a healthy person capable of donating more frequently unable to donate. This is highly an acceptable scenario in a situation where there is an allover lack of blood in most of the blood banks and hospitals. Therefore an effective system for the purpose of next blood donation intimation has been elaborated in this research article. The intimation for next blood donation is achieved through k-means clustering and Pearson correlation along with deep belief network and decision tree approaches. The experimental results have been satisfactory for the accuracy of the provided suggestion.**

***Keywords:*** *K Means Clustering, Pearson Correlation, Deep Belief Networks and Decision Tree*

## I. INTRODUCTION

Every person's health is the foundation of a fulfilling existence. One of the important components that must be kept in excellent shape to appreciate the many phases and phases in life is one's health. There are several viruses, germs, and other effects that can be harmful to one's health. There has been a continuing battle to maintain an efficient approach for sustaining an individual's health. Individual health is maintained by the use of numerous ways in the medical industry.

The medical platform has been crucial in reaching many goals, such as illness eradication and trauma treatment milestones. The majority of the difficulties that the human race faces are caused by lifestyle choices and misfortunes such as accidents and other traumas. This has by far the greatest influence on an individual, who can be quite difficult to cure of a medical standpoint. Medical advancements have been beneficial in achieving and improving this individual's lifestyle.

Individuals undergoing treatment for blood-related issues or who are victims of accidents frequently lose a significant amount of blood. Blood is an essential component of our bodies that cannot be manufactured synthetically at this time. Therefore, blood is a vital resource that may be used to save such people and their lives. The sole supply of blood comes from a healthy person who donates blood to successfully provide the injured individual or victim. Therefore healthy individuals are strongly encouraged to donate blood, as there is always a shortage of blood in blood banks and hospitals across the world.

A healthy person can give blood, which can be utilized to save the life of a patient suffering from numerous illnesses or an accident victim. The individual can only donate a certain amount of blood-based on his or her health and only after a certain length of time has passed. Please provide sufficient time between donations to guarantee that the individual has created The Lost blood and may donate again without causing harm to the donor. However, because every person's body is different, there are varying timeframes that are restricted to people regarding the recovery of given blood.

Because blood is a precious resource, and healthy persons can make it more easily and quickly, having to wait for the required waiting period might result in a variety of losses. This decrease in blood bank efficiency may result in much less blood coming in for use by sufferers. Therefore, there is a need for an efficient system that can notify and person about the next blood donation time that is tailored to their body type. For this aim, a group of researchers was discovered and investigated in this piece, which aided us in reaching our conclusion for such a system.

This literature survey paper dedicates section 2 for analysis of past work as a literature survey, and finally, section 3 concludes the paper with traces of future enhancement.

The second section of this research article focuses on describing the current researches in the field. The implementation is described in full in Section 3 of the suggested approach. The achieved result is discussed in Section 4 named as results and discussions. Finally, part 5 brings this study piece to a close, as well as the scope of future improvements.

## II LITERATURE SURVEY

R. Ali explains that there should be a regulated system in place to handle all blood donors and transfusion operations. [1] All data should be saved in a centralized database. The suggested document establishes uncontrolled blood banks and parallel marketplaces to promote awareness and boost confidence. To prevent uncontrolled donation procedures from infected donors, donor data should be registered in the system, as well as in blood banks and campaigns. Thus, the framework assists blood banks and donors in saving patients' lives through a regulated approach.

A. Meiappane states the number of blood donors varies per country. Several blood donation camps offer but cannot give enough assistance to the needy[2]. The suggested framework provided a solution by developing the Haversine Mathematical formula and by benefactor information of the donors. When a donor donates blood, the donor's information is automatically removed for the next three months. Following that, the program employs a haversine mathematical procedure to locate the nearest available donor. The major goal of the article is to deliver the necessary facts in as little time as possible.

M. Fahim explains that the health industry is rapidly expanding, although the services supplied to citizens are inadequate[3]. The compatibility of blood groups with the patient and the donor is a very difficult issue in the medical field. As a result, the researcher created an android-based blood donation application as a way to link the requester and donor at any time and from any location. The author's major goal is to present to the requester the number of owners available in his area. The architecture of the system is separated into three sections: volunteer blood donors, cloud computing, and blood requester.

D. Domingos emphasizes that to meet the world's blood needs, drastic measures must be used. [4] Many nations use various techniques to increase the number of blood donors, such as South Africa's Club 25, Ghana's national blood policy, which governs the blood donation and transfusion process while also attempting to educate, inspire, attract and retain young blood donors. The suggested article used the 'gamification' notion to develop an application called 'Blood Hero,' in which users are rewarded for performing social deeds after giving blood.

F. Alharbi states that healthcare systems rely on blood donation to preserve lives, hence blood donation is the primary source of blood supply in many nations. [5] Formal paraphrase After studying the donor's blood cycle, the suggested paper offers information technology solutions. This research is carried out under the Central Blood Donation Management System (CBDMS), which is suggested as a network of interconnected systems. It was the first e-management system for managing blood donations. The introduction of CBDMS reduces the time necessary for blood donation simply by reducing the information requested from donors. The suggested solution improves the effectiveness of the blood donation management system.

L. Evdochim explains that by implementing flow rate estimation in existing blood donation systems employing different force sensors. The PPG-based blood flow rate estimation method is used in the suggested article. PPG is an abbreviation for photoplethysmograph, which is a picture of blood flow fluctuations in the human body. [6] Formalized paraphrase Two conditions must be met for the photoplethysmogram hypothesis to work. The first requires that the examined material or specimen be separated from the surrounding environment by optical qualities, while the second requires that the specimen be in motion or, in other words, that its local distribution develops over time.

A. Mantari's goal is to provide a framework for participants' willingness to donate blood. Also, to learn more about blood donation and its advantages [7]. Participants as blood donors can be encouraged by presenting this information, and this strategy is used to raise awareness before blood donation campaigns on university campuses. The questionnaire in the proposed article is created with the use of data gathered from the public's perspectives on blood donation. The purpose of this is to raise awareness before blood donation.

Blood transfusion is important for human life in perilous conditions, according to P.KIRCI. Thus, the number of potential donors and blood donation probability may be estimated using a variety of machine learning algorithms [8]. Machine learning algorithms aid blood transfusion procedures by utilizing datasets. Machine learning is used to detect patterns in enormous amounts of data. Algorithms execute the learning and machine tasks by utilizing the supplied data. Based on performance, one of the most effective classification algorithms was compared on the blood transfusion data set.

M. Kamalesh states that when blood drives are held, blood donation is referred to as donating blood. One of the most vital and significant aspects of human life is blood. Obtaining blood in an emergency is extremely difficult [9]. The suggested study proposes a simple method for locating a local donor who can aid in an emergency case. If a person needs blood, he or she can submit a request via a website or mobile application, and the request will be forwarded to someone who matches the matching blood group in a local area.

L. Evdochim describes that photoplethysmogram is one of the most successful procedures for detecting cardiovascular tissues. This approach may also be used to effectively and accurately identify the rate of blood flow. The blood flow rate may be used efficiently to detect many illnesses and other irregularities in the heart [10]. The technique's accuracy is heavily dependent on the wavelength and optical reflectance of the light source. Therefore, the photoplethysmogram approach was used in this study to obtain very accurate blood flow rate identification, which may be effective in diagnosing different inconsistencies and arrhythmias with ease.

## III PROPOSED METHODOLOGY

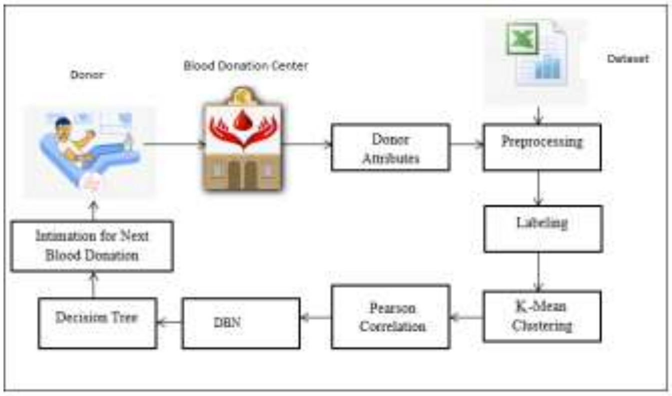


Figure 1: System overview for the Blood Donation Intomation system

The presented technique for determination of blood donation interval through the use of deep belief networks has been illustrated in the figure 1. The methodology has been attained through the utilization of stepwise procedure given below.

*Step 1: Dataset collection, preprocessing and Labeling* – In the first step of the proposed methodology for blood donation interval estimation the data of the donor is collected through to the blood bank employee. The interactive user interface design through swing has been utilized to collect the attributes of the donor.

The data set consisting of various parameters of previous donors in various blood banks across the world have been provided as an input to this system by downloading it from the URL https://www.kaggle.com/bonastreyair/predicting-blood-analysis?select=blood-train.csv.

The data set is read into the system by the Java code through the utilization of the jxl API. This API is useful in providing the workbook format of the data set as input which is transformed into a double dimension list for easier interfacing with the proposed system.

The input data set may contain various missing attributes or irregular data that needs to be conditioned before providing it as an input to the system. The preprocessing procedure significantly eliminates all these redundancies and selects only the useful attributes for our approach such as volume donated number of donations, months elapsed since first donation and months elapsed since last donation. These attributes are effectively labeled in the list which is provided as an input to the next step of the procedure for clustering.

*Step 2: K Means Clustering* – This procedure takes the labeled data set attributes as an input along with the user attributes enter by the employer for performing the operation. The preprocessed and labeled data set is clustered according to the k-means clustering algorithm which groups together similar data for achieving the clusters. The process for cluster formation has been stipulated in the steps below.

*Distance Evaluation* – The Euclidean distance is being used to evaluate the distance between the user input and the data set attributes. This is necessary for evaluation of each and every row in the data set for their relevance with the user input. The distance is evaluated for each of the attributes for each of the rows in the data set are appended to the end of their respective rows. The distance is evaluated through the use of the equation 1 given below.

$$ED = \sqrt{\sum(ATi - ATj)^2} \quad \_\_\_\_\_\_\_\_ \ (1)$$

Where,
ED=Euclidian Distance
$A_{Ti}$=Attribute at index i
$A_{Tj}$= Attribute at index j

*Centroid Estimation* – The row distances attend the previous step are utilized to find the average row distance calculating the mean. The list is then effectively sorted in the ascending order of the row distances. The data points are than estimated for K number of integers randomly from the sorted list. These data points are then used for extracting the respective row distances pertaining to these data points.

These data points are used to extract the respective row distances which will now be referred to as the centroid. This centroid value is highly useful as it provides the reference for achieving the clusters. The row distances along with the average row distance is used for calculation of the cluster boundaries. This is done by addition and subtraction off the row distance from the average distance to achieve the maximum and minimum values for the boundaries.

The grabbed average distance and the row distance are considered as the boundaries of the cluster that will be achieved by the addition and subtraction of these values from one another. These achieved boundaries will be useful in the next step for the assimilation of the clusters.

*Cluster Formation* – Cluster formation - the boundaries along with the centroids achieved in the previous step are useful for limiting the entries in the list and clubbing them together in the form of clusters. The minimum and maximum values can be effectively used for clamping the attributes in the data set together to obtain the cluster list. There are k clusters now which are effective in segregating the attributes that can be provided to the next step of the approach for the purpose of estimating the correlation.

*Step 3: Pearson Correlation* – The estimation of correlation in our system has been achieved through the utilization of the Pearson correlation approach. The input to this step is the cluster list obtained previously through the use of k means clustering. These clusters are based on the user input provided by the employer based on the attributes of the user which are also used as in input in this step for correlation. The correlation estimation uses X and y arrays for the purpose of correlation calculation. Where in the value of x is considered as the independent variable array whereas the dependent variable array is referred to as y.

The dependant variable value or x in our system is considered as the input attributes of the donor provided by the employee. Whereas the dependent variable value is considered as the clusters obtained through the previous step. These values are then subjected to the equation to given below to provide the correlation value for a particular row of attributes.

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{(x^2 - \frac{\sum x^2}{n})}\sqrt{(y^2 - \frac{\sum y^2}{n})}} \quad \text{\_\_\_\_\_\_\_\_ } \mathbf{(2)}$$

Where,
x is the user input attributes
y is the attributes extracted from the dataset
n is the total number of entries
r= correlation value in between -1 to +1.

This process is depicted in the algorithm 1.

**ALGORITHM 1: Cluster Correlation Estimation**

```
//Input : K-Means Cluster List KC_L , User Input list X[ ]
//Output: Cluster Correlation List  CCR_L
clusterCorrelationEstimation(KC_L, X)
Start
CCR_L =∅
  for i=0 to size of KC_L
      S_L= KC_L [i] [ Single Cluster]
      TMP =∅ [Temporary List]
      for j=0 to size of S_L
          R= S_L[j]
          Y[ ]→ R
          CR_VAL=pearsonCorrelation(X,Y)
          R=R+ CR_VAL
          TMP= TMP+R
      end for
   CCR_L = CCR_L + TMP
  end for
 return CCR_L
Stop
```

This procedure for correlation measurement is performed iteratively for each and every row in the achieved clusters. These values of correlation range from -1 to 1. The correlation values of each of the rows of the cluster are 1st added and then divided by the number of rows in that particular cluster. This results in an average correlation value for a particular cluster which is considered as the cluster correlation value. This is performed for all of the clusters achieved and then these clusters are sorted in descending order of these correlation values. The top three clusters are then provided to the next step of deep belief networks further processing.

*Step 4: Deep Belief Neural Networks* – The correlation values of the clusters achieved in the previous step are utilized as an input to this step of the procedure. The deep belief networks are provided the top three clusters based on the achieve the correlation values having maximum correlation with the user input. These clusters are subjected to input layer and hidden layer formation to achieve the output layer. For this purpose 4 random weights are assigned along with two bias weights with the target values for determining the output layer as mentioned in the equations 3 and 4 given below.

X= (AT1* W1) + (AT2*W2) +B1 _____ (3)

$$H_{LV} = \frac{1}{(1+\exp(-X))} \quad \text{\_\_\_\_\_\_\_ (4)}$$

Where,

W1, W2 - Random weights,
(Other random weights set is also used like {W1, W2, W3, W4, B1, B2})

AT1, AT2 - Attributes for which prediction probability is estimated

$H_{LV}$ – Hidden layer

The hidden layers have the ReLu activation function which provides the estimation of the hidden layer through to the given input weights along with the bias weights. This attains hidden layer values from the equations given above for the assessment of the outer layer and extraction of the error probability.

The values of the output layer achieved from the above equations are combined with the target values in the equation 4 to achieve the error probability rate. The target value is here are considered as 0.99 and 0.01.

$$Error\ Probability = \sum \frac{1}{2} (T_0 - O_L)^2 \text{----- (4)}$$

Where,
T = Target Values
$O_L$ = Output Layer Values

The value of error probability achieved through this equation is appended to the end of the respective rows of the cluster to form a double dimensional list. This is combined with the actual input values and provided to the next step for the classification and suggestion generation.

*Step 5: Decision Tree* – The decision tree is a highly effective classification approach that can completely classify a particular input into classification labels through the use of if - then rules. The error probability values along with the provided input are given to this step of the methodology for the purpose of classification. This value is combined with the total donated volume to achieve the error correction value. This error correction value is subjected to classification for the purpose of determining the donation interval of the donor. This suggestion is provided to the donor for the purpose of their next visit to the blood bank.

## IV RESULTS AND DISCUSSIONS

The proposed methodology for the purpose of achieving intimation for the next blood donation through the use of deep belief networks have been realized through the Java programming language. The development environment of NetBeans has been utilized for writing the code of this methodology. The machine for development of this approach was powered by an Intel core i5 processor assisted through a 4GB of Ram and 500 GB of hard disk storage.

The present approach needs to be effectively evaluated for the presence of any errors to determine the execution accuracy of the intimation estimation for next blood donation. This is crucial as the suggestion of the blood donation is a very critical scenario that needs to be highly precise and accurate to reduce any mishaps. Therefore the performance evaluation has been effectively derived through the realization of the accuracy attained by the methodology.

Mean reciprocal rank will be utilized for achieving the performance metrics of the proposed methodology. As the intimation for the next blood donation is being provided in the form of a suggestion this suggestion needs to be evaluated for its relative accuracy. This will be done through to a particular rank given to the system by a human user. The humans are are highly accurate identifiers of this outcome due to the inherent nature of analysis achieved by them.

A number of 10 users have experimented on the proposed system for the evaluation of the suggestion provided by the system for the next blood donation. These users have provided varying input for the attributes utilized in this methodology and the accuracy of the resultant suggestion is scrutinized and rank is provided from 1 to 6.

The reciprocal of this rank is utilized for the purpose of achieving a value between 0 and 1 where the rank of 1 is a highly accurate suggestion and a rank of 6 is a completely inaccurate suggestion. The mean of the resultant values is achieved and the outcomes of this experimentation have been tabulated in the table 1 below.

| User | Donation Interval Evaluation Rank |
|---|---|
| 1 | 1 |
| 2 | 1 |
| 3 | 0.5 |
| 4 | 0.5 |
| 5 | 1 |
| 6 | 1 |
| 7 | 1 |
| 8 | 0.5 |
| 9 | 1 |
| 10 | 1 |
| MRR | 0.85 |

Table 1: User Mean Reciprocal Rank (MRR) for blood donation interval suggestion

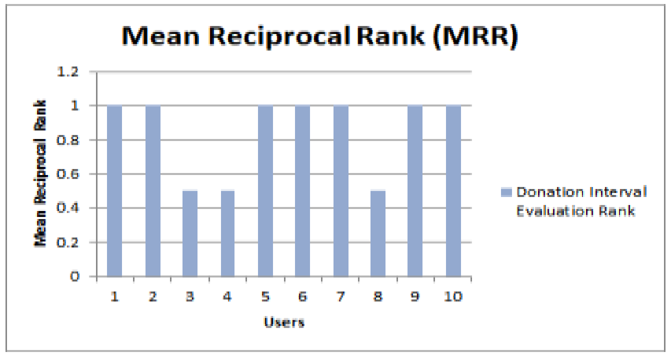


Figure 2: Graphical Representation of the MRR Values

The figure 2 describes the bar graph achieved for or the main reciprocal rank given to the proposed methodology. The presented approach for intimation of next blood donation achieves mean reciprocal ratio of 0.85 which has been consistent through a number of users. This cements the execution accuracy of our approach for the purpose of providing accurate suggestions. This also proves that the deep belief networks and decision tree approaches have been implemented accurately in the proposed methodology for intimation for next blood donation.

## V. CONCLUSION AND FUTURE SCOPE

The presented technique for intimation of the next blood donation has been achieved through deep belief networks and decision tree. There is a need for an effective approach that can significantly reduce the blood donation interval and provide an accurate donation interval based on the user attributes. This is achieved by the proposed methodology through taking the input attributes of the donor along with a data set for the purpose of preprocessing and labeling. This input values are effectively preprocessed and label and then provided to the k-means clustering module for the purpose of clustering the similar inputs. The clustering approach employs the use of Euclidean distance to perform the distance evaluation and eventually cluster formation. The achieved clusters are then provided to the next step of the approach for correlation estimation through Pearson correlation. The Pearson correlation approach determines the correlation between the inputs attributes and the data set clusters achieved previously. This correlation values for each of the clusters are sorted in the descending order and the top three clusters with the maximum amount of correlation are provided as a input to the deep belief networks. The deep belief networks effectively utilizes these clusters to determine the hidden values which are clubbed together with the target values to achieve the output layer as an error probability value. The user input and error probabilities are used for error correction which are then correlated with the user input to provide the accurate suggestion for blood donation interval. The experimental results achieved true evaluations have indicated a satisfactory accuracy for the suggestion.

The future research direction can be focused on deploying this approach in the form a mobile application for the easier prediction of the next date of blood donation.

## REFERENCES

[1] R. S. Ali, T. F. Hafez, A. B. Ali and N. Abd-Alsabour, "Blood bag: A web application to manage all blood donation and transfusion processes," 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, 2017, pp. 2125-2130, doi: 10.1109/WiSPNET.2017.8300136.

[2] A. Meiappane, K. Logavignesh, R. Prasanna and T. Sakthivel, "D'WORLD: Blood Donation App Using Android," 2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN), Pondicherry, India, 2019, pp. 1-5, doi: 10.1109/ICSCAN.2019.8878830.

[3] M. Fahim, H. I. Cebe, J. Rasheed and F. Kiani, "mHealth: Blood donation application using android smartphone," 2016 Sixth International Conference on Digital Information and Communication Technology and its Applications (DICTAP), Konya, 2016, pp. 35-38, doi: 10.1109/DICTAP.2016.7543997.

[4] D. C. L. Domingos, L. F. S. G. Lima, T. F. Messias, J. V. L. Feijó, A. A. R. Diniz and H. B. Soares, "Blood hero: An application for encouraging the blood donation by applying gamification," 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, 2016, pp. 5624-5627, doi: 10.1109/EMBC.2016.7592002.

[5] F. Alharbi, "Progression towards an e-Management Centralized Blood Donation System in Saudi Arabia," 2019 International Conference on Advances in the Emerging Computing Technologies (AECT), Al Madinah Al Munawwarah, Saudi Arabia, 2020, pp. 1-5, doi: 10.1109/AECT47998.2020.9194178.

[6] L. Evdochim, A. E. Zhdanov, V. I. Borisov and L. G. Dorosinsky, "Blood Mixers for Transfusion Therapy: Method of Blood Flow Velocity Determination Based on Photoplethysmogram," 2020 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT), Yekaterinburg, Russia, 2020, pp. 0175-0179, doi: 10.1109/USBEREIT48449.2020.9117670.

[7] A. Alva Mantari, B. A. Guzmán, A. J. Lipa Cueva, G. Delgado-Rivera, O. P. Trinidad Quiñonez and A. Roman-Gonzalez, "Raise of Awareness for Blood Donation Campaigns on University Campuses in Lima, Peru," 2018 Congreso Internacional de Innovación y Tendencias en Ingeniería (CONIITI), Bogota, 2018, pp. 1-4, doi: 10.1109/CONIITI.2018.8587061.

[8] P. KIRCI, S. AKTAS and B. SEVINC, "Analyzing Blood Donation probabilities and number of possible donors," 2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), Ankara, Turkey, 2020, pp. 1-4, doi: 10.1109/HORA49412.2020.9152872.

[9] M. D. Kamalesh, A. M. J., Y. Felix, D. S. and M. Prasad, "Automation of Blood Donation by Data Integration Using Data Mining," 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), Tirunelveli, India, 2020, pp. 944-948, doi: 10.1109/ICOEI48184.2020.9143010.

L. Evdochim, A. E. Zhdanov, V. I. Borisov, D. Dobrescu and L. G. Dorosinsky, "Blood Mixers for Transfusion Therapy: Blood Flow Estimation Based on PPG Technique," 2020 13th International Conference on Communications (COMM), Bucharest, Romania, 2020, pp. 49-53, doi: 10.1109/COMM48946.2020.9142037.

*****