

Design of Data Mining Techniques for Online Blood Bank Management by CNN Model

I. Jeena Jacob¹, P. Ebby Darney²

¹Department of Computer Science and Engineering, GITAM University, Bangalore, India

²Department of Electrical and Electronics Engineering, RajaRajeswari College of Engineering, Bangalore, India

E-mail: ¹jeenajacob2016@gmail.com, ²darney.pebby@gmail.com

Abstract

A blood bank is the organisation responsible for storing blood to transfuse it to the patients in need. The primary goal of a blood bank is to be reliable and ensure that patients get the relevant non-toxic blood to avoid transfusion-related complications since blood is a critical medicinal resource. It is difficult for the blood banks to offer high levels of precision, dependability, and automation in the blood storage and transfusion process if blood bank administration includes many human processes. This research framework is proposing to maintain blood bank records using CNN model classification method. In the pre-processing of CNN method, the datasets are tokenized and set the donor's eligibility. It will make it easier for regular blood donors to donate regularly to charitable people and organizations. A few machine learning techniques offer the automated website updation. Jupyter note book has been used to analyze the dataset of blood donors using decision trees, neural networks, and von Bays techniques. The proposed method operates online through a website. Moreover, the donor's eligibility status with gender, body mass index, blood pressure level, and frequency of blood donations is also maintained. Finally, the comparison of different machine learning algorithms with the suggested framework is tabulated.

Keywords: Machine learning, online blood bank, CNN data mining, automatic website update

1. Introduction

In a blood bank, donated blood is stored and maintained for the use of blood transfusions in the future. When we talk about "blood bank," we usually mean a section of a hospital where blood products are stored and tested as needed (to reduce the risk of transfusion-related adverse events). The term "collection centre" may also apply to a hospital, which does the collection of blood. Anaemia and iron deficiency patients may benefit from whole blood or RBC transfusions to increase their blood oxygen saturation [1-3].

To locate information, individuals need to access vast libraries of data stored both online and offline. In today's era, computerised techniques can do it more efficiently than human power. Large datasets in a particular application sometimes need data mining to uncover previously undiscovered nuggets of information [4-6].



Figure 1. Screenshot of an online blood bank management system

The human search for information in big archives has become more challenging as data sizes have exploded. In order to locate the required information from data warehouses and repositories, such as corporate organisations with internet-based or offline data, a computer-based technique has been used. A screenshot of an online blood bank management system is shown in Figure 1 on its donor search window.

Android-based Blood Donation System stores, processes, retrieves, and analyses data related to blood bank administration and inventory management. This project's goal is to keep track of all the blood donors and the various blood types available at each blood bank so that things can be better organized. This document serves as a model for creating new documents in the future.

A whole blood transfusion is requisite for situations requiring a lot of blood, such as accident victims or patients having extensive surgery. For reasons of safety, online registration is required. Only authorised and registered hospitals will be able to use the website to refrain from fraudulent act. It may also restrict access to just those who are truly in need of it. A donor or a member of the hospital's administration may be the user. When a patient visits the website, the exact location of the hospital, as well as the blood type needed has to be entered [7-11]. The blood bank and donor information is stored in a database. Blood banks are divided up by district to make things simpler when it comes to keeping track of donations. By using geolocation, the hospital's address is transformed into geocoordinates. In order to calculate the distance between a blood bank's stored address and a hospital's address, the haversine algorithm is employed [12, 13].

2. Organization of the Research

The rest of the research paper is divided into the following sections: Section 3 summarizes previous research on methods for managing online blood banks. Section 4 describes the machine learning algorithm's suggested approach for an online blood bank management system. Section 5

summarizes the findings and discusses them. The last part summarizes the study findings and makes recommendations for further work.

3. Preliminaries

A management information system was developed by G. Satyanarayana Reddy and colleagues to assist managers in making decisions across all kinds of organizations. When it comes to data mining, this MIS is all about getting the most out of the data that's already out there [14].

Red Blood Cell Counting was used to diagnose various illnesses by Ivana D et al. via the Health Care Applications. In this study, researchers demonstrated an automated method for counting red blood cells (RBCs) from pictures using data mining methods such as segmentation, equalisation, and K-means clustering. Predictions may also be made for illnesses linked to sickness and a low RBC count, such as leukemia [15].

There is an Android Blood Donor life-saving Cloud Computing application developed by T. Hilda Jenifha et al. to connect all blood donors to aid and manage a blood transfusion service and create a database to store data on blood supplies in each region of the city. The database contains information on donors who used supply chain networks and the RVD Algorithm. However, there are restrictions since supply chain networks are more complicated than other methods [16].

In order for the mining system to be effective, it should be able to work with minimum user input, to determine which information would be most useful to the user and to extract that useful knowledge from a huge medical dataset. The system should be able to provide usable knowledge and extract manageable collections [17].

Researchers Javed Akhtar Khan and colleagues have developed a new rural blood bank administration system based on Cloud Computing. This method enables individuals to give blood to anybody who needs it at any time or in any circumstance. In addition, the seeker may contact the donor in an emergency by dialing the emergency number. Since it just keeps track of Blood Bank statistics, this system doesn't pay attention to how donors respond when they have to donate blood [18].

Data mining methods were used by Ramachandran et colleagues to categorize blood donors. Using an inefficient method, the research hopes to locate frequent contributors who donate freely from their convenient time. A J48 type classification algorithm was used on a typical blood group donor's dataset to conduct the study. The J48 model is chosen based on the models' accuracy in making classifications. Blood donors were analyzed based on their sex, blood type, weight, and age for the study. Using data mining methods and choosing the appropriate implementation tools for the domain area, it is possible to succeed. Despite this, the study only utilized a few minor characteristics to categorize the blood donors and thus the categorization type is a little hazy [19].

Research Gap

Despite a large number of prospective blood donors, only a tiny proportion of the Indian population donates blood. On the other hand, there has been a significant rise in the need for blood. Potential blood donors have been discouraged from donating since the procedures bring forth the lack of appropriate architectural flow. Few observations from the past methods;

1. Due to the database's large size, it's almost impossible to locate the different fields listed above. Blood banks aren't in the business of keeping an inventory.
2. Multiple queries are inefficiently handled by systems, resulting in incorrect results.
3. When blood banks run out of supplies, they can't assist those in need.

4. Any metropolitan city blood group donor registry does not exist.
5. Blood waste is not being prevented since there is no mechanism in place.

4. Methodologies

The CNN model is used to classify the donor's eligibility status and accurate prediction of need in the hospital. Figure 2 explores the architecture of the proposed work.

4.1 Pre-processing for CNN

Using data preprocessing, outliers that might skew the results of data mining can be identified and eliminated, as well as missing values can be filled in and noisy data can be detected and removed. The data is also collected from filtered fields, with appropriate techniques indicated to alleviate missing data symptoms and redundant data included. The knowledge-based analysis is used to preprocess and modify the dataset. Several machine learning and data mining methods, including deep learning, use the Java platform [20, 21].

4.1.1 Tokenization

Encryption is often used by businesses to safeguard data while it is in transit or at rest. In addition to encryption, there are additional data protection methods, such as tokenization. It is possible to tokenize sensitive information by transforming it to a non-disclosable value, such as a random number. While it's not required, a token vault may store both plain text and its corresponding token in a safe location if one is available. Tokenization isn't widely utilized since the procedure produces an unintelligible, non-reversible token. If a vault-less tokenization technique is employed, tokenization may be irreversible.

4.1.2 Normalization of dataset

Blood groupings may be assigned to separate tables. Given the fact that your blood type will always match that of the donor, it is probably not needed to place the blood group on the blood bag table. Therefore, the donor id is not needed in the transaction database since it will always be the same as the donor of the blood bag in the transaction table.

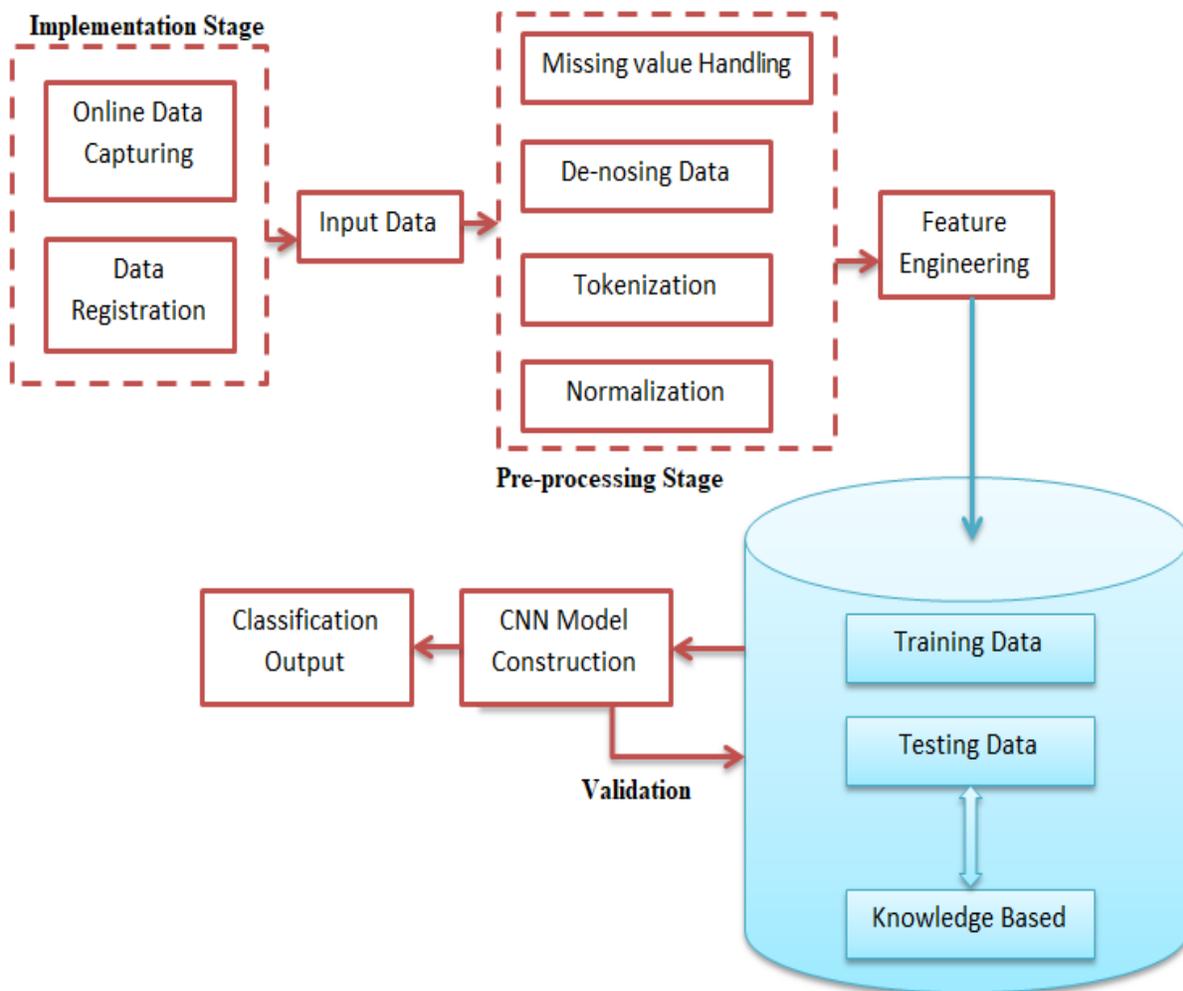


Figure 2. Architecture of the proposed model

4.2 Classification

Different data mining methods were analyzed and compared, and classification was chosen as the basis for this model's development. Predicting the value of a single nominal variable using established groups and known values for the other components is one of the most important jobs. For the purpose of categorizing blood donor health status, a classification model CNN-based algorithm is being developed [23-26].

4.3 Implementation stage

To put it in another way: Implementation helps to put everything prepared into action and moves the project from planning to service delivery. The blood bank website system is implemented using several languages.

4.3.1 HTML

The primary purpose of this program is to build web pages. The web browser's primary function is to read documents as web pages, but it is also possible to incorporate scripts written in a variety of languages, such as JavaScript, that affect how web pages behave.

4.3.2 PHP

An HTML scripting language that adds functionality isn't possible with basic HTML. PHP enables to gather processes and use data to generate the required output, since it was originally intended to build dynamic web pages. Because PHP works well with databases, queries, and other capabilities, it's a great choice for web development. Many types of data, information, and images are stored in MySQL. It's also quite easy to go there from any part of the globe.

4.3.3 Java script

Interactive websites and online apps are created with this programming language. Since JavaScript interacts well with HTML source code, web writers may add dynamic content to their sites. Blood bank website systems are implemented using programs.

4.3.4 Control panel

The website graphical control panel is meant to make site administration easier. The CNN Panels on the facade manage every element of the facility. Commercial web hosting providers may utilise CNN Panel's proprietary software, which is provided via cPanel. Consequently, the business does not provide reduced prices to individuals. Despite this, non-profit organisations like educational institutions and charities, on the other hand, may seek a license to use it for free or at a reduced cost [27-29].

4.3.5 Output

Using the numerous CSS page layouts available, highly efficient and flexible websites may be created. Web pages may now contain Ajax capabilities as well. Ajax allows the fast and easy development of web applications.

5. Results Discussion

Blood Transfusion Service Center Data Set is used for testing and training. The multivariate characteristics have been used in this dataset for real attributes. The classification associates tasks with 4 attributes. Table 1 shows the proposed CNN model classification instance error rate and positive rate.

Table 1. Classification instance error rate

Model	No. of Total Instances	Incorrect Classified Instances	Correctly Classified Instances	Correctly Classified rate in percentage	Classification error rate in percentage
Naïve Bayes	1500	114	1386	92.4%	7.6%
Decision Tree	1500	75	1425	95%	5%
Proposed CNN	1500	37	1463	97.53%	2.467%

The name, type, and measurement unit of the variable will all be updated as a result of this modification. This listing's order corresponds to the database's row numbering order.

R is for Rhetorical (Recency-months since last donation)

F (Frequency-total number of donations)

M is a good example of this (Monetary-total blood donated in c.c.)

T is a binary variable that represents the amount of time since the initial contribution.

Table 2. Performance analysis on various algorithms

Model	Accuracy	F-Measure	Recall	Precision	Computation Time (Sec)
Naïve Bayes	82.4%	78%	79%	80%	4.21
Decision Tree	86%	82%	83%	85.2%	3.78
Proposed CNN	98.23%	93%	91.45%	94.87%	6.05

The computation time required by this suggested CNN model is somewhat longer than that required by other conventional methods. This suggested method achieves excellent results in the other performance measures as well.

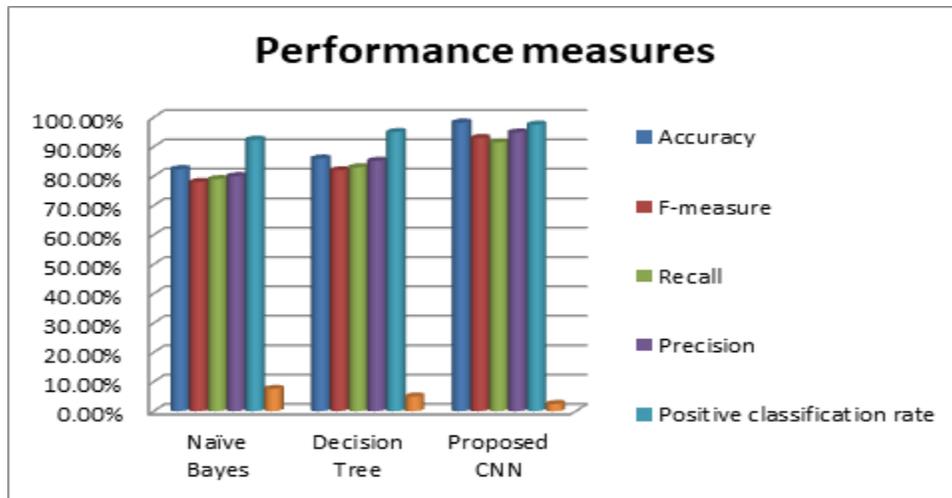


Figure 3. Overall performance measures

F-Measure, accuracy, recall, precision, and computing time were utilised to assess the classification models' overall performance. The confusion matrix formula was used to analyse the data in table 2. Figure 3 shows overall performance measures of various models. The confusion matrix shows which classifications are accurate and which are wrong. The right classifications are indicated by True Positive (TP) and True Negative (TN), while the erroneous classifications are indicated as "False Positive" (FP) and "False Negative" (FN), respectively.

6. Conclusion

This research article introduces data mining and the proposed CNN method to find intriguing patterns and create important connections between variables in a large dataset. Many industries rely on data mining to sift through mountains of data and find the nuggets of information

they require. Medical data such as gender, BMI, blood pressure, and frequency of blood donations are only a few examples of the types of information that must be kept in every area to retain records. The future work will focus on improving these models further for the integration with blood benefactor frameworks, including novel perception techniques. This research may one day be followed in the entire country's blood banks. The collected data from the area may be subjected to Big Data analysis. As a result, a particular domain name may be assigned to the blood bank. In addition, the blood bank's website may allow blood donors to request blood units online by providing them with a login id and password.

References

- [1] Sungeetha, Akey, and Rajesh Sharma. "Design an Early Detection and Classification for Diabetic Retinopathy by Deep Feature Extraction based Convolution Neural Network." *Journal of Trends in Computer Science and Smart technology (TCSST)* 3, no. 02 (2021): 81-94.
- [2] Subramanian, D. Venkata, R. Sugumar, K. Maneesh, V. Manoj, and A. Venkat Kishore. "Effective Record Search and Update Using Design Patterns-A Case Study of Blood Bank Mobile Application." In *International conference on Computer Networks, Big data and IoT*, pp. 39-47. Springer, Cham, 2019.
- [3] Haoxiang, Wang, and S. Smys. "Big Data Analysis and Perturbation using Data Mining Algorithm." *Journal of Soft Computing Paradigm (JSCP)* 3, no. 01 (2021): 19-28.
- [4] Singh, Vernika, Raju Shanmugam, and Saatvik Awasthi. "Preventing Fake Accounts on Social Media Using Face Recognition Based on Convolutional Neural Network." In *Sustainable Communication Networks and Application*, pp. 227-241. Springer, Singapore, 2021.

- [5] Bashar, Abul. "Survey on evolving deep learning neural network architectures." *Journal of Artificial Intelligence* 1, no. 02 (2019): 73-82.
- [6] Aiswaryadevi, V. J., S. Kiruthika, G. S. Priyanka, N. Nataraj, and M. S. Sruthi. "Effective Multimodal Opinion Mining Framework Using Ensemble Learning Technique for Disease Risk Prediction." In *Inventive Computation and Information Technologies*, pp. 925-933. Springer, Singapore, 2021.
- [7] Manoharan, J. Samuel. "A Novel User Layer Cloud Security Model based on Chaotic Arnold Transformation using Fingerprint Biometric Traits." *Journal of Innovative Image Processing (JIIP)* 3, no. 01 (2021): 36-51.
- [8] Silva, Jesus, Jesús García Guliany, Lissette Hernandez, Rafael Portillo, Noel Varela, Hugo Hernández Palma, Osman Redondo Bilbao, and Lesbia Valero. "Data mining and neural networks to determine the financial market prediction." In *International Conference on Communication, Computing and Electronics Systems*, pp. 361-368. Springer, Singapore, 2020.
- [9] Smys, S., Joy Iong Zong Chen, and Subarna Shakya. "Survey on Neural Network Architectures with Deep Learning." *Journal of Soft Computing Paradigm (JSCP)* 2, no. 03 (2020): 186-194.
- [10] Urooj, Aksa, Md Tabrez Nafis, and Mobin Ahmad. "An Analytical and Comparative Study of Hospital Re-admissions in Digital Health Care." In *Computer Networks, Big Data and IoT*, pp. 717-733. Springer, Singapore, 2021.
- [11] Tripathi, Milan. "Analysis of Convolutional Neural Network based Image Classification Techniques." *Journal of Innovative Image Processing (JIIP)* 3, no. 02 (2021): 100-117.
- [12] Shyam Sundaram and Santhanam T; 'Real Time Blood donor management using Dash boards based on Data mining models', *International Journal of Computing issues*, Vol.8, Issue5, No.2, September 2011.

- [13] Vijayakumar, T. "Comparative study of capsule neural network in various applications." *Journal of Artificial Intelligence* 1, no. 01 (2019): 19-27.
- [14] G. Satyanaryana Reddy et al; 'Management Information System to help Managers for providing decision making in any organization'.
- [15] Ivana D. Radojevic et al. 'Total coliforms and data mining as a tool in water quality monitoring', *African Journal of Microbiology Research*, Vol. 6(10), 16 March 2012.
- [16] Jenipha, T. Hilda, and R. Bachiyalakshmi. "Android Blood Donor Life Saving Application in cloud computing." *Biometrics and Bioinformatics* 6.1(2014):34-36.
- [17] M. Kay, J. Santos and M. Takane, "mHealth: New horizons for health through mobile technologies", World Health Organization, 2012
- [18] Khan, Javed Akhtar, and M.R Alony. "A New Concept of Blood Bank Management System Using Cloud Computing for Rural Area (INDIA)." *International Journal of Electrical, Electronics and Computer Engineering*, 4(1), p.20
- [19] P. Ramachandran et al. 'Classifying blood donors using data mining techniques'; *IJCST*, Vol. I, Issue 1, February 2011.
- [20] F. Zabihi et al. 'Rule Extraction for Blood donors with fuzzy sequential pattern mining'; *The Journal of mathematics and Computer Science*; Vol. II, Issue No. I, 2011.
- [21] Joe, Mr C. Vijesh, and Jennifer S. Raj. "Location-based Orientation Context Dependent Recommender System for Users." *Journal of trends in Computer Science and Smart technology (TCSST)* 3, no. 01 (2021): 14-23.
- [22] Vikram Singh and Sapna Nagpal; 'Interactive Knowledge discovery in Blood Transfusion Data Set'; *VSRD International Journal of Computer Science and Information Technology*; Vol. I, Issue 8, 2011.
- [23] Suma, V., and Shavige Malleshwara Hills. "Data Mining based Prediction of Demand in Indian Market for Refurbished Electronics." *Journal of Soft Computing Paradigm (JSCP)* 2, no. 02 (2020): 101-110.

- [24] Adam, Edriss Eisa Babikir. "Deep Learning based NLP Techniques In Text to Speech Synthesis for Communication Recognition." *Journal of Soft Computing Paradigm (JSCP)* 2, no. 04 (2020): 209-215.
- [25] Wen-ChanLee and Bor-Wen Cheng; 'An Intelligent system for improving performance of blood donation', *Journal of Quality*, Vol. 18, Issue No. II, 2011.
- [26] Manoharan, J. Samuel. "Capsule Network Algorithm for Performance Optimization of Text Classification." *Journal of Soft Computing Paradigm (JSCP)* 3, no. 01 (2021): 1-9.
- [27] Karunakaran, P., and Yasir Babiker Hamdan. "Early Prediction of Autism Spectrum Disorder by Computational Approaches to fMRI Analysis with Early Learning Technique." *Journal of Artificial Intelligence* 2, no. 04 (2020): 207-216.
- [28] Vikram Singh and Sapna Nagpal; 'Interactive Knowledge discovery in Blood Transfusion Data Set'; *VSRD International Journal of Computer Science and Information Technology*; Vol. I, Issue 8, 2011.
- [29] Hamdan, Yasir Babiker. "Faultless Decision Making for False Information in Online: A Systematic Approach." *Journal of Soft Computing Paradigm (JSCP)* 2, no. 04 (2020): 226-235.

Author's biography

I. Jeena Jacob is working as a Professor in the Computer Science and Engineering department at GITAM University, Bangalore, India. She actively participates in the development of the research field by conducting international conferences, workshops and seminars. She has published many articles in refereed journals. She has guest edited an issue for *International Journal of Mobile Learning and Organisation*. Her research interests include mobile learning and computing.

P. Ebby Darney is working as an Associate Professor in the Department of Electrical and Electronics Engineering, RajaRajeswari College of Engineering, Bangalore, India. His area of research includes Image Processing, Artificial Intelligence, Control Systems, Radio Networks, and cloud computing.