

An Intelligent Blood Management System Using OPTICS Clustering and Real-Time Donor Tracking

Rosiya A.S.¹, Niranjana S.², Krisha C.R.³, Preethi Monika K.⁴

^{1,2,3}Students, ⁴ Assistant Professor, Department of Artificial Intelligence And Data Science,
NPR College of Engineering and Technology, Dindigul, India.

E-mail: ¹rosiyaas920822243043@nprcolleges.org, ²niranjanas920822243037@nprcolleges.org,

³krishacr920822243026@nprcolleges.org, ⁴preethimonikak@nprcolleges.org

Abstract

Effective blood donor selection involves the efficient prioritization of donors, compatibility checking, inventory validation, and spatial location-based donor identification. This research proposes an intelligent blood management system including four-class donor priority classification, eligibility and blood group validation, inventory first validation, spatial clustering using OPTICS algorithm, geographic prioritization, and donor tracking after acceptance. For donor priority classification, Blood Transfusion Service Center dataset comprising 748 donor data and four numerical attributes-recency, frequency, monetary value of blood volume, and time-has been used. The developed classification algorithm shows 0.95 precision, 0.93 recall, and 0.94 F1-score while the confusion matrix analysis shows that there is an effective segregation between the four classes of donor priorities having 91.37% classification accuracy. Furthermore, the results of multiclass ROC analysis confirm the discriminatory capability of the model having AUC values ranging from 0.95 to 0.99.

Keywords: Donor-Priority Classification, OPTICS Spatial Clustering, Blood-Group Compatibility Filtering, Inventory-First Verification, Geographical Donor Prioritization, Post-Acceptance Tracking, Intelligent Blood Management.

1. Introduction

Availability of safe and compatible blood in a timely manner is required for emergency situations, surgeries, traumas, and diseases that require regular blood transfusions. Existing blood management systems mainly rely on static donor database, manual communication, and blood group or proximity searches, which may lead to donor latency, repeated alerts, and poor management during emergency cases. With increased use of machine learning and location-based services, there is an option to create smarter donor management systems.

ML can support in making predictions regarding the response and eligibility of the donor, whereas spatial analysis can be used in order to make the process of geographically locating donors more effective. Specifically, the algorithm of OPTICS (Ordering Points to Identify the Clustering Structure) helps in identifying density-based donor clusters without the need to define the number of clusters, which makes this algorithm suitable for the case of dynamically distributed donors. Nevertheless, current algorithms concentrate only on donor identification or registration and do not have much visibility once the donor responds to the request.

To address these limitations, the research proposes an advanced and intelligent blood management system that combines donor prediction, donor eligibility filtering, blood group compatibility testing, OPTICS-based clustering of donors, inventory prioritization, notification, and post-acceptance monitoring. The process starts by checking the availability of compatible blood inventory and, upon the unavailability of adequate stocks, initiates donor solicitation. After that, eligible and available donors are clustered and prioritized based on their spatial and operational characteristics, and automatic notifications are generated. Next, the donors who are eligible and available are clustered and prioritized based on operational and spatial factors.

2. Literature Review

Digital transformation has become more and more prevalent in the fields of blood donation and blood banking operations to facilitate donor identification and accessibility. Data-driven approaches using machine learning techniques have been studied for predicting donor willingness and possible donor prediction, indicating that data-driven methods can be effective tools for better donor prediction and sufficient donor pool maintenance [1], [4], [7], [8]. The researches in this field mostly concern themselves with donor classification and prediction of donor willingness

and possible donors' identification who are more willing to donate their blood. But these works concentrate only on donor prediction, not considering the whole process of the emergency request for blood donation.

Blood management systems using mobile technology and web-based platforms have been created to facilitate the exchange of information among donors, blood banks, and the healthcare community [6], [9], [10]. In addition, such systems have features that include donor registration, blood bank records, availability of blood, and processes related to donations and transfusions. Cloud-based solutions help in accessing and exchanging information at centralized locations through blood bank operations [5]. The above technologies make it easier to manage donor information, but the main drawback is that most of them do not involve spatial donor groupings and monitoring of donor after acceptance.

Donor suitability is also affected by factors related to health and time related to recurrent donations of blood. Several studies have looked at the link between iron deficiency and blood donation, demonstrating that donor health and intervals of donation should be considered when determining donor suitability [2]. Other research on the effect and experience of regular blood donation underscores the importance of taking care of donor suitability [3]. Therefore, including these constraints in a computer-based algorithm for selecting donors is crucial in order to avoid unsuitable donor selection.

Although existing researches on ML-based donor prediction, blood management systems for mobile and web applications, centralizing blood bank system, and donor eligibility issues [1]-[10] have been carried out, very little effort has been invested in the integration of these approaches with density-based donor clustering and tracking of the donor right after acceptance. Traditional donor recognition approaches normally stop once a donor who meets the requirement has been found. The question of whether the accepted donor is actually end route to the site of donation remains a concern during emergencies where time is of essence. Thus, an integrated approach is necessary that involves donor eligibility filter, donor clustering, prioritization of blood stock, donor alerting, and real-time tracking of donor.

3. Proposed Methodology

The proposed system consists of a modular process for donor search, blood request handling, spatial grouping, and donor monitoring after acceptance. The whole process can be summarized in Figure 1, where the interrelation between donor data processing, donor eligibility validation, blood type matching, inventory validation, donor clustering using the OPTICS algorithm, donor selection, notification, and real-time tracking is presented. This process starts with the validation of the availability of the needed blood unit before starting the donor search process. Once there is a need for more donors, they are selected based on their classification and location.

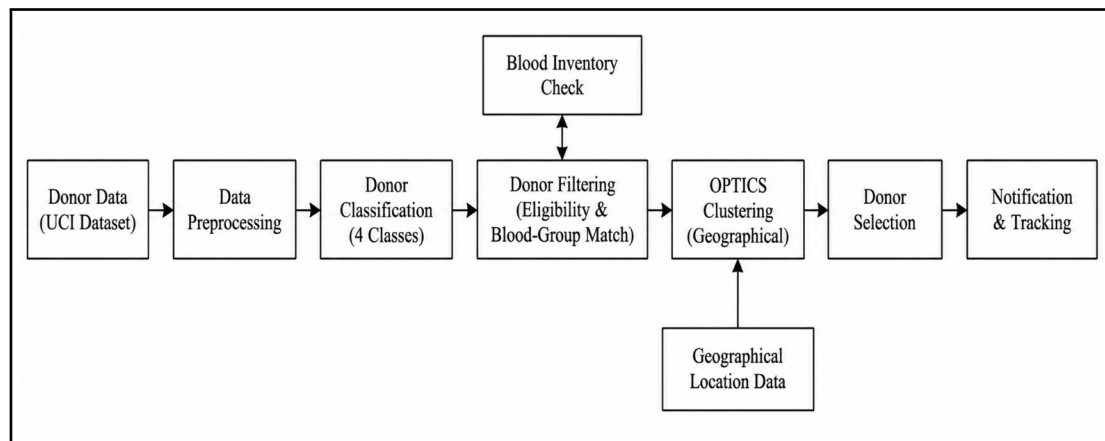


Figure 1. Workflow of the proposed intelligent blood donor selection and management System

As illustrated in Figure 1, the methodology starts from donor data collection and pre-processing, followed by donor feature extraction and classification. After that, the extracted donor data are merged with blood-group compatibility, availability, and eligibility constraints. In the case where the necessary stock of blood is not available, the OPTICS algorithm is used to arrange the appropriate donors based on their location. The donors thus identified are contacted and accepted donors are directed to the post-acceptance donor tracking process.

3.1 Dataset Description and Classification

The Blood Transfusion Service Center dataset downloaded from the UCI Machine Learning Repository is used to perform the donor response classification [11]. The Blood Transfusion Service Center dataset has 748 donor entries with four input variables, which include recency, frequency, monetary blood volume, and time. The output variable states whether the

donor gave blood, and thus, this represents a binary classification task. The dataset properties are provided in Table 1.

Table 1. Dataset characteristics of the blood transfusion service center dataset

Dataset Characteristic	Description
Dataset Name	Blood Transfusion Service Center
Dataset Source	UCI Machine Learning Repository [11]
Number of Instances	748
Number of Input Features	4
Input Features	Recency, Frequency, Monetary, Time
Feature Type	Numerical
Target Variable	Donor Classification
Classification Type	Multiclass Classification
Number of Classes	4
Class 0	Low Donor Priority
Class 1	Moderate Donor Priority
Class 2	High Donor Priority
Class 3	Very High Donor Priority
Missing Values	None
Data Format	Tabular
Prediction Objective	Donor Priority Classification

The data is pre-processed such that the donor dataset is divided into four priority levels according to their historical donation behavior. The classes are represented as numbers ranging from 0 to 3, with Class 0 being a donor with low priority, Class 1 being a donor with moderate priority, Class 2 being a donor with high priority, and Class 3 being a donor with very high priority. Four-class representation helps the classification algorithm differentiate between levels of donor suitability and priority instead of deciding whether a donor is suitable or not. Evaluation is done based on precision, recall, F1-score, ROC-AUC, and the four-class confusion matrix.

3.2 Data Pre-processing

The donor data are initially analyzed for any duplicate, incomplete, and inconsistent entries prior to model building. The four numerical attributes – Recency, Frequency, Monetary and Time – are selected as input variables and converted to a uniform numerical format. As these attributes could have varying ranges, feature scaling is performed to eliminate bias associated with differences in scales when training the model. The scaled value of each feature is determined by

$$z_i = \frac{x_i - \mu}{\sigma} \quad (1)$$

where x_i denotes the original feature value, μ represents the mean of the corresponding feature, σ represents its standard deviation, and z_i denotes the standardized feature value.

After the normalization process, the pre-processed data is sorted into input feature and label target data sets. The target data is coded based on the four classification labels provided in Table 1 and ranging from Class 0 to Class 3. This data set is later split into training and test data sets to facilitate modeling and performance assessment. This pre-processing technique is consistently applied to both the training and test data sets for the classification process.

3.3 Donor Response Classification

The pre-processed donor data is used by the classification algorithm, which classifies each donor based on the four priority categories provided in Table 1. The learning algorithm learns from the pattern between the characteristics of historical donations and the priority class associated with the donor. The classification algorithm has four outputs and uses labels starting with Class 0 for donors with lower priority and up to Class 3 for very high priority.

For each test data record, the model produces a predicted class label that is then compared against the actual class label to analyse classification effectiveness. The predictions generated in this way are then used to compute the confusion matrix with four classes along with other metrics including precision, recall, F1 score, and ROC-AUC. These metrics help evaluate the model from two different perspectives – one by looking at the predictive accuracy per class and the other by measuring discrimination power. The classification result is further used in the

donor selection module.

3.4 Blood-Group Compatibility and Donor Eligibility Filtering

Following the donor classification procedure, the records of the candidates are filtered based on compatibility in terms of blood group and eligibility. The blood group required by the person/organization is compared with the blood groups of the donors to check their compatibility. The donors who do not fulfil the eligibility requirements of the particular blood group are then excluded from the procedure.

This predicted priority class from the classification phase is kept for further ranking of candidates. Therefore, the filtering approach does not solely depend on the outcome from the classification phase, as compatibility of blood groups, eligibility, availability, and priority of the donors are considered prior to carrying out the spatial clustering. This ensures that incompatible and unavailable donors do not appear in the cluster generated using OPTICS.

3.5 Blood Inventory Verification

The proposed workflow follows an inventory first principle to check whether the blood demand can be met from the current inventory before any donor selection takes place. Whenever there is a demand for blood, the type of blood and the number of units required are matched with the current inventory of compatible blood. In case the inventory contains the sufficient amount, the demand is satisfied from the existing inventory and no donor selection becomes necessary. In case of insufficient availability in the inventory, the workflow activates the donor selection process from among the eligible donors. This will ensure that only the deficit blood is selected from the donors. The inventory decision process can be expressed as.

$$D_I = \begin{cases} 1, & Q_a \geq Q_r \\ 0, & Q_a < Q_r \end{cases} \quad (2)$$

where Q_a is the available compatible blood quantity and Q_r is the requested quantity. Here, $D_I=1$ indicates sufficient inventory, while $D_I=0$ activates donor selection.

3.6 OPTICS-based Donor Clustering

When the inventory verification identifies an unmet requirement, the potential donors who have been filtered out will then undergo Ordering Points to Identify the Clustering Structure (OPTICS). The objective here is to arrange the eligible donors depending on their geographical location and classify the donors that can be found in dense geographic areas. As opposed to partition-based clustering techniques that set a priori the number of clusters needed, OPTICS gives an order with respect to its density and reachability properties.

This algorithm assesses the neighborhood structure for each donor based on spatial distance and density. The reachability distance between each candidate point is calculated as

$$RD_{p,q} = \max\{CD_p, d_{p,q}\} \quad (3)$$

where $CD(p)$ denotes the core distance of donor p , and $d(p,q)$ represents the spatial distance between donors p and q . This supports in the clustering process by determining how near the donor can reach from another density connected donor. The result OPTICS order is utilized in determining clusters of donors that are meaningful, which will then be passed to the donor selection process. Clustering therefore helps to narrow the search space of candidates and yet maintain geographically meaningful donor clusters for the blood requirement.

3.7 Geographical Distance and Donor Prioritization

Following the clustering using the OPTICS algorithm, the selected donors within the related clusters are then sorted based on the geographical closeness to the donation site. This is done by calculating the minimum possible geographical distance between the donor and the target destination. For two locations represented by latitude 1,2 and longitude 1,2, the Haversine formulation is used:

$$d = 2R \sin^{-1} \left(\sqrt{\sin^2 \left(\frac{\Delta\phi}{2} \right) \cos\phi_1 \cos\phi_2 \sin^2 \left(\frac{\Delta\lambda}{2} \right)} \right) \quad (3)$$

where R denotes the Earth's radius, represents the difference in latitude, and represents the difference in longitude. The derived distance becomes an important factor for the ranking of

donors in the formed clusters. Those who have their locations near the location where donation is needed get preference in the selection process, but donor-priority classes and donor qualifications remain the other two factors in the process of donor selection.

3.8 Donor Notification and Response Management

After selecting donors based on the priority list, the donors will be notified for targeted notification according to their priority, eligibility, compatibility in terms of blood group, availability, and geographical proximity. The notification stage involves notifying the prioritized donors of the need for blood and documenting their reactions for the particular request. The donors, who react positively to the particular request are retained for the next stage of tracking.

The response status for every donor who receives notification is kept updated with regard to the present status of the recruitment process. This is necessary to ensure that the system is able to differentiate between pending responses, accepted responses, and unavailability, and to avoid reprocessing the same candidate once the request is already processed. The accepted donor's details are then handed on to the post-acceptance tracking process.

3.9 Real-time Post-acceptance Donor Tracking

The donation request being accepted by a donor, the donor file then moves on to the tracking of donor after acceptance phase. This phase entails tracking the geographical location of the donor through their available location until they reach the set donation location. This phase tracks how far away the donor is from the set location. The donor's location is compared with the destination location at periodic intervals.

The tracking process ends when the donor arrives at the specified location for donation or otherwise ends. The donor's status is updated accordingly, showing that the tracking phase has ended. With the connection between donor acceptance and tracking, there is continuity from donor selection, acceptance of request, monitoring of movement, and finally arrival at the location for donation.

4. Results and Discussion

4.1 Donor-priority Classification Performance

The donor-priority classification approach was tested based on the metrics of precision, recall, and F1-score, which show how well the model is able to differentiate between four identified donor-priority classes. As shown in Figure 2, the classification model produced the following results: the precision value was 0.95, recall value was 0.93, and F1-score was 0.94. Based on the results obtained, it is possible to say that the values of the three measures are quite similar, indicating a balanced behavior of the model.

The precision value indicates that most of the data points classified into a certain donor priority fall under the correct category, whereas the value of recall shows that a large number of records from their respective categories have been detected by the model. The F1 score is a measure of these two properties and further suggests the consistency in the classification. These findings confirm the utility of the predicted donor priority class as a decision-making factor along with other factors like eligibility, blood group compatibility, availability, and location in the next step of donor selection.

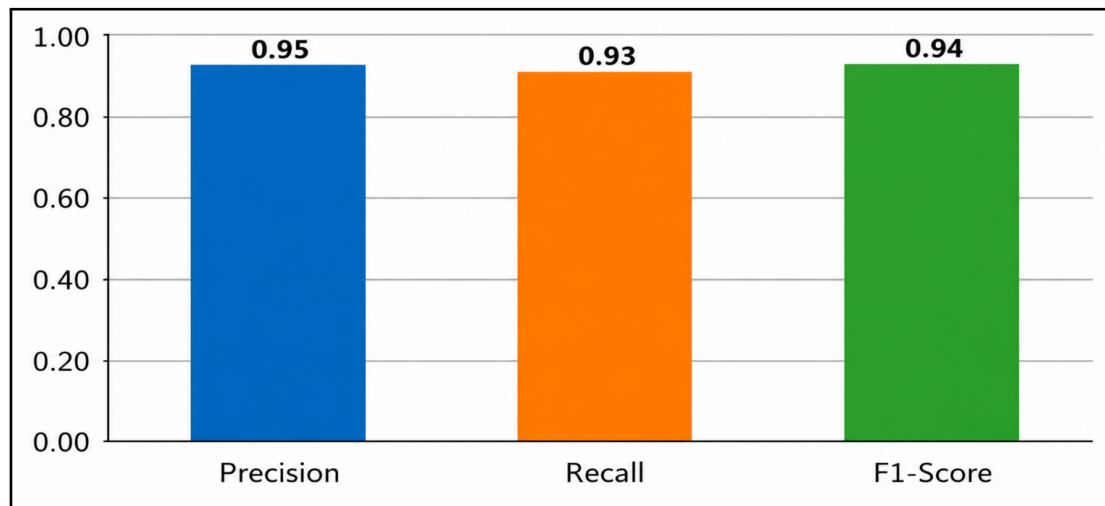


Figure 2. Performance metric comparison of the donor-priority classification model

4.2 Confusion Matrix Analysis

Donor-priority classification prediction characteristics on a class-by-class basis are described through the use of the confusion matrix illustrated in Figure 3. Four types of donor priorities have been considered in the construction of the confusion matrix, that is, Class 0 for

Low Donor Priority, Class 1 for Moderate Donor Priority, Class 2 for High Donor Priority, and Class 3 for Very High Donor Priority.

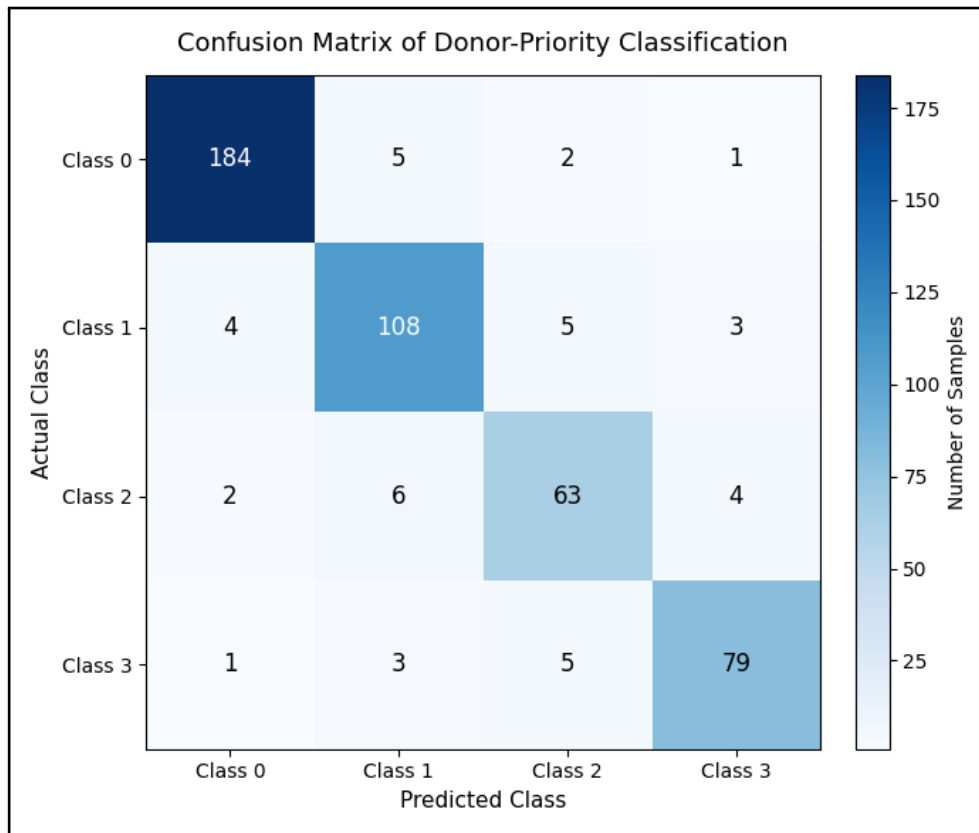


Figure 3. Confusion matrix of the four-class donor-priority classification model

The matrix consists of 475 instances where the number of correctly classified data is 434, and hence the classification accuracy of 91.37% is achieved. Class 0 has the highest number of correct predictions, which is 184 instances. Classes 1, 2, and 3 have 108, 63, and 79 correct predictions, respectively. From the dominating diagonal, it is clear that the model has been able to differentiate between most of the donor instances by their priority classes.

The number of samples of Class 2 classified as belonging to Class 1 or Class 3 is six and four, respectively. It demonstrates that some donors with high priority have features that resemble those of moderate and very high priority donors. In addition, there is no error between Classes 0 and 1 and between Classes 2 and 3. Therefore, the errors are found more between priority level and adjacent levels but not throughout all levels.

4.3 Training and Validation Accuracy Analysis

The training and validation accuracy graphs presented in Figure 4 show how the donor-priority classification model learns throughout the training process. Training accuracy shows an increase as we move through the initial training epochs, implying that the model is able to learn the dependencies between the input donor features and priority classes allocated. Validation accuracy also shows an increase in a similar pattern and eventually becomes stable.

During the later epochs, both the curves remain relatively stable, which means that the model has become stable with respect to its learning. The validation curve has not degraded significantly at the time when the training accuracy is improving, which means that overfitting is not very severe. Stabilization means that the classification representation learned by the model has become stable with respect to its learning.

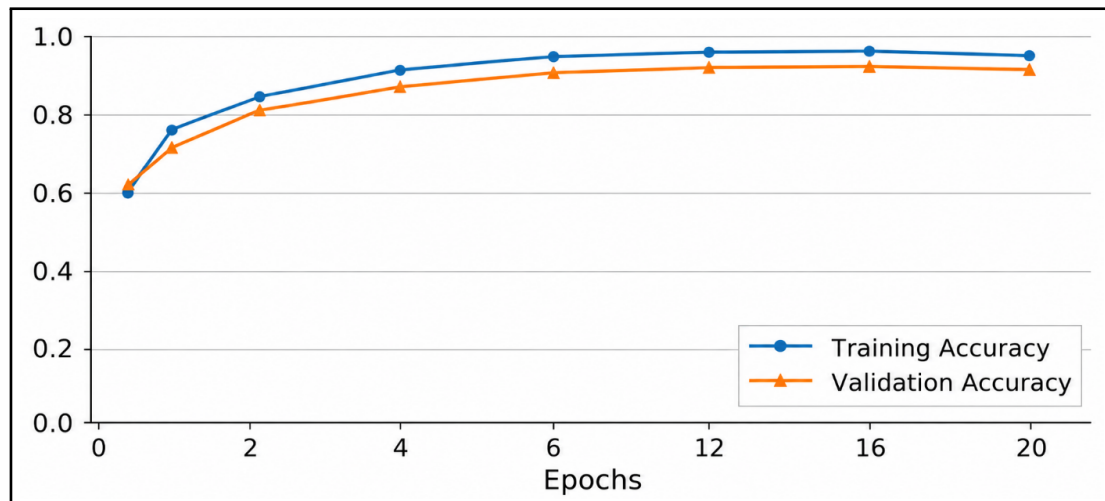


Figure 4. Training and validation accuracy of the donor-priority classification model

4.4 ROC-AUC Analysis

The ROC curve analysis illustrated by figure 5 assesses the discriminative ability of the donor-priority classifier among the four classes that have been specified. The ROC curves for Class 0, Class 1, Class 2, and Class 3 are all far from the reference random guess curve line, suggesting that there is successful discrimination among the donor-priority classes.

The class-wise AUC scores are 0.99 for Class 0, 0.97 for Class 1, 0.95 for Class 2, and 0.96 for Class 3. Class 0 shows the best discrimination, while Class 2 shows the least discrimination, based on the AUC scores. However, all four ROC curves show much better

discrimination compared to the random-guess AUC score of 0.50.

The similarity of AUC values from all four donor priority classes indicates that the performance of the classification model is consistent in terms of discrimination and not biased towards one particular donor priority class. The ROC analysis thus also confirms that the donor features obtained have a good discriminating ability for the four donor priority classes.

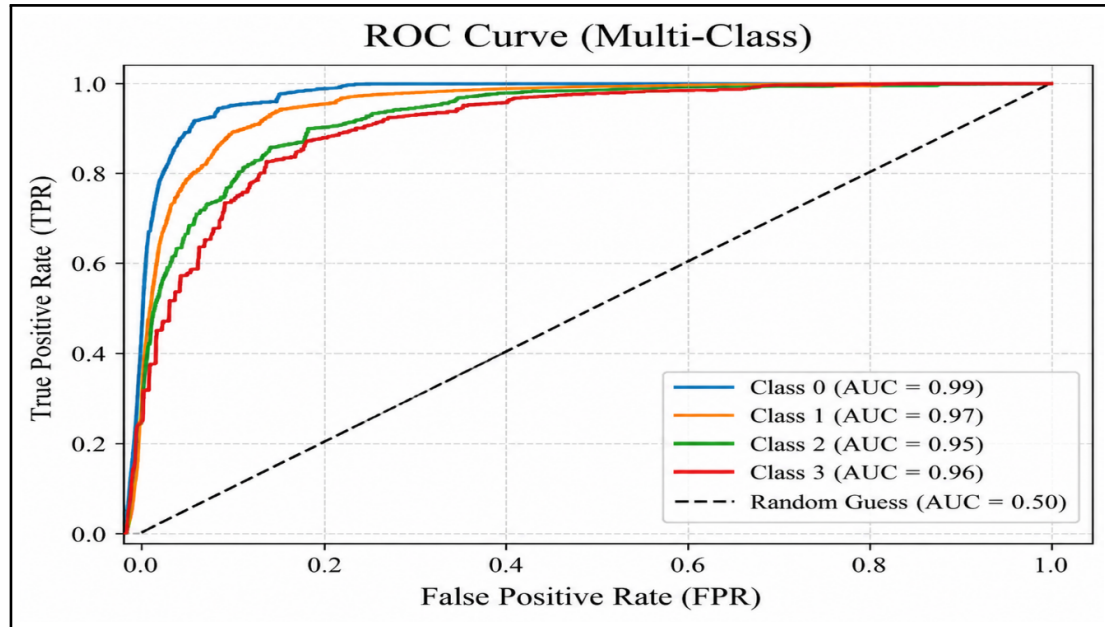


Figure 5. Multiclass ROC curves for donor-priority classification

4.5 Integrated Donor Selection and Management Analysis

Integration of classification results with donor eligibility, blood group compatibility, availability, and blood stock status help in the entire process of donor selection. The predicted priority category of donors acts as a supportive criterion and not as a criterion for donor selection on its own. In case there is sufficient blood availability in the inventory which is also compatible, then no donor recruitment takes place. In case there is insufficient availability of blood, then the eligible and compatible donors are sent to the clustering step using OPTICS.

The workflow is then followed by a geographic closeness of the donor selection prior to notification. The accepted donors are thereafter moved to the post-notification tracking phase, thereby ensuring smooth transition from donor selection to movement tracking. Thus, the suggested method integrates donor priority classification, eligibility and compatibility screening, inventory validation, spatial clustering, geographic prioritization, notification, and

post-notification tracking into a sequence of blood request management.

5. Conclusion and Future Work

The proposed intelligent blood management system encompasses a sequence of decisions such as donor-priority classification, eligibility and blood-group compatibility filtering, inventory validation, OPTICS clustering, geographical priority, notification, and acceptance tracking. With such an integrated approach, identification and prioritization of potential donors will be completed considering factors such as availability of blood types and geographical proximity. By adopting an inventory-first approach, the number of unnecessary donors can be reduced, while the geographical priority approach will facilitate the targeting of candidates. The future work may involve expanding the approach by employing bigger and more diversified donor databases, real-time streams of blood bank inventories, demand forecasting, and dynamic clustering methods. More sophisticated approaches involving ensembles and DL may also be explored in order to analyze the intricate behavior of the donors. Data privacy protection and multi-center validation can be additionally included for increased security and reliability.

References

- [1] Wu, Hong-yun, Zheng-gang Li, Xin-kai Sun, Wei-min Bai, An-di Wang, Yu-chi Ma, Ren-hua Diao et al. "Predicting Willingness to Donate Blood Based on Machine Learning: Two Blood Donor Recruitments During COVID-19 Outbreaks." *Scientific Reports*, 2022, vol. 12, no. 1: 19165.
- [2] Mantadakis, Elpis, Paraskevi Panagopoulou, Eftychia Kontekaki, Zoe Bezirgiannidou, and Georges Martinis. "Iron Deficiency and Blood Donation: Links, Risks and Management." *Journal of Blood Medicine* 2022, vol. 13: 775-786.
- [3] Makowicz, Dawid, Renata Dziubaszewska, Katarzyna Lisowicz, and Natalia Makowicz. "Impact of Regular Blood Donation on The Human Body; Donors' Perspective. Donors' Opinion on Side Effects of Regular Blood Donation on Human Body." *Journal of Transfusion Medicine and Hemostasis*, 2022, vol. 15, no. 2: 133-140.
- [4] Wahono, Hermanto. "Prediksi Calon Pendonor Darah Potensial Dengan Algoritma Naïve Bayes, K-Nearest Neighbors Dan Decision Tree C4. 5." *JURIKOM (Jurnal Riset Komputer)*, 2020, vol. 7, no. 1: 7-14.

- [5] Chaudhari, Shreyas Anil, Shrutika Subhash Walekar, Khushboo Ashok Ruparel, and Vrushali Milind Pandagale. "A Secure Cloud Computing Based Framework for The Blood Bank." In International Conference on Smart City and Emerging Technology (ICSCET), IEEE, 2018, 1-7.
- [6] Sugianto, A., and M. Zundi. "Design and Build a Mobile-Based Blood Donor Application At PMI Bandung Regency." KOPERTIP: Jurnal Ilmiah Manajemen Informatika dan Komputer, 2017, vol. 1, no. 1: 11-18.
- [7] Atmaja, Ketut Jaya, Ida Bagus Gede Anandita, and Ni Kadek Ceryna Dewi. "Penerapan Data Mining Untuk Memprediksi Potensi Pendonor Darah Menjadi Pendonor Tetap Menggunakan Metode Decision Tree C. 45." S@ CIES, 2017, vol. 7, no. 2: 101-108.
- [8] Boonyanusith, Wijai, and Phongchai Jittamai. "Blood Donor Classification using Neural Network and Decision Tree Techniques." In Proceedings of the World Congress on Engineering and Computer Science, 2012, vol. 1: 499-503.
- [9] Ali, Rehab S., Tamer F. Hafez, Ali Badawey Ali, and Nadia Abd-Alsabour. "Blood Bag: A Web Application to Manage All Blood Donation and Transfusion Processes." In International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), IEEE, 2017, 2125-2130.
- [10] Kaur, Manvir, Nahida Nazir, Navneet Kaur, Syed Faraz Ali, Chirag Agarwal, Ujjwal Dubey, Varun Gupta, Abid Sarwar, Manik Rakhra, and Omdev Dahiya. "A Web-Based Blood Bank System for Managing Records of Donors and Receipts." In International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), IEEE, 2022, 459-464.
- [11] Yeh, I., Blood Transfusion Service Center [Dataset]. UCI Machine Learning Repository, 2008. <https://doi.org/10.24432/C5GS39>.