

Uso de Aprendizado de Máquina para Predizer o Comportamento do Doador de Sangue: Uma Revisão da Literatura

Jairo Marciano Silva¹, Ana Maria Monteiro¹

¹Centro Universitário Campo Limpo Paulista - UNIFACCAMP
jmarciano1971@gmail.com, anammont.alu@gmail.com

Abstract. *Hemotherapy Services, commonly called Hemotherapy Units, often have trouble recruiting blood donors in times of crisis or in emergency situations like the recent COVID-19 pandemic. In addition, in these services there is a constant concern to keep blood stocks at safe and acceptable levels. This article presents a systematic review of the literature on the use of machine learning techniques that help to obtain a better understanding of the factors that influence the future behavior of the donor to be able to predict it with greater precision, in order to define strategies to retain and increase the number of active donors. We reviewed 17 articles, selected from 171 articles initially retrieved from 5 indexed databases. The article also discusses 10 papers considered most relevant for understanding the behavior of donors.*

Resumo. *Serviços de Hemoterapia, chamados comumente de Unidades Hemoterápicas, muitas vezes tem problemas para convocar doadores de sangue em épocas de crise ou em situações de emergência como a recente pandemia do COVID-19. Além disso, nesses serviços existe uma preocupação constante em os estoques de sangue em patamares seguros e aceitáveis. Este artigo apresenta uma revisão sistemática da literatura sobre o uso de técnicas de aprendizado de máquina que ajudem a obter uma melhor compreensão sobre os fatores que influenciam o comportamento futuro do doador para poder prevê-lo com maior precisão e assim definir estratégias para convocar e aumentar o número de doadores ativos. Foram revisados 17 artigos, selecionados dos 171 artigos recuperados inicialmente de 5 bases indexadas. O artigo, também, discorre sobre os 10 trabalhos considerados mais relevantes para o entendimento do comportamento dos doadores.*

1. Introdução

A transfusão sanguínea é fundamental para o tratamento de pacientes em casos de emergências graves causadas, por exemplo, por acidentes e também é um recurso terapêutico de extrema importância para pacientes com anemias falciformes, hemofilia, leucemia aguda, entre outras, que necessitam de transfusões regulares [Magalhães et al. 2017]. Outro aspecto importante que vale a pena ressaltar, é a natureza perecível do sangue total que possui um prazo de validade de aproximadamente de 42 dias a partir da

data de coleta. Após a coleta, o sangue total é frequentemente dividido em concentrado de hemácias, concentrado de plaquetas randômicas e plasma com prazo de validade de 42, 5 e 365 dias respectivamente.

Recentemente o sangue e seus componentes foram adicionados à lista de medicamentos essenciais da Organização Mundial de Saúde¹ (OMS), mas sua disponibilidade é insatisfatória nos países de renda *per capita* baixa ou média, levando a mortalidade e morbidade que poderiam ser evitadas [Roberts *et al.* 2019].

Segundo a OMS, em 2013, ocorreram aproximadamente 112.5 milhões de doações de sangue no mundo todo. Desse total, somente 54% são efetuadas em países de baixa e média renda, sendo que estes concentram 82% da população mundial [WHO 2017]. No Brasil, considerando dados de 2017 do Ministério da Saúde, ocorreram aproximadamente 3.4 milhões de doações de sangue e foram realizadas 2.8 milhões de transfusões [Ministério da Saúde 2019].

A motivação principal para doação de sangue é devido a um comportamento pró-social, a uma ideologia sem vantagens óbvias para o doador além da satisfação pessoal. Outros fatores que levam as pessoas a doar é já ter precisado de sangue para si ou para pessoas próximas, algumas pessoas só vão atrás de exames sorológicos e outras são motivadas devido a estarem envolvidas com grupos que participam de campanhas de doação [Gillespie e Hillyer 2002].

Os doadores de sangue podem ser classificados como:

- **Doador de Primeira vez (New Volunteer Donor – NVD):** doador de sangue voluntário não remunerado, que nunca doou sangue antes;
- **Doador Esporádico (Irregular Volunteer Donor – IVD):** doador de sangue voluntário não remunerado que doou sangue no passado, mas não cumpre os critérios de um doador regular;
- **Doador de Repetição (Regular Volunteer Donor – RVD):** é um doador de sangue voluntário não remunerado que doa sangue regularmente, conforme uma frequência pré-definida.

Neste artigo realizamos uma Revisão Sistemática de Literatura (RSL), com o objetivo de identificar técnicas de aprendizado de máquina que ajudem a obter uma melhor compreensão sobre os fatores ou variáveis que influenciam o comportamento futuro do doador. O conhecimento desses fatores permitirá prever esse comportamento com maior precisão e assim definir estratégias para convocar e aumentar o número de doadores com potencial para se tornarem RVD.

Nesta revisão, utilizamos a metodologia apresentada em [Kitchenham 2007], que possui três fases diferentes: planejamento, execução e publicação dos resultados. A fase de planejamento tem como finalidade identificar a motivação para a execução de uma RSL. Na sequência é definido o protocolo da revisão, especificando as questões de pesquisa, estratégia de busca, fontes de pesquisa, *string* de busca, critérios de seleção (inclusão e exclusão) e critérios de qualidade. Para Kitchenham (2007), a não utilização de um protocolo para seleção de estudos primários e a posterior análise desses estudos, poderia gerar intercorrências no processo, devido a direcionamentos tendenciosos dados à

¹ <https://www.who.int/>

pesquisa para atender as expectativas do pesquisador, tornando os resultados suspeitos ou duplicados.

Na fase de execução, são identificados e selecionados os estudos primários utilizando-se das definições da estratégia de busca e dos critérios de seleção e qualidade. Ainda nessa fase realiza-se a extração e sintetização dos dados. Já na fase de publicação é responsável por descrever, divulgar e avaliar os resultados relatados.

O restante deste artigo está estruturado da seguinte maneira: a Seção 2 apresenta a etapa de planejamento; a Seção 3 a execução da revisão; a Seção 4 sintetiza os principais resultados obtidos descrevendo brevemente 10 artigos selecionados e detalha 2 artigos considerados mais relevantes; e, por fim, a Seção 5 apresenta as conclusões do artigo.

2. Planejamento

A RSL teve como base os artigos publicados entre os anos 2010 à 2021 nas bases *PUBMED*², *ACM*³, *IEEE*⁴, *SPRINGER*⁵ e *Google Scholar*⁶.

A questão principal que a RSL busca identificar na literatura científica é: “Quais são os estudos primários sobre a utilização de métodos de aprendizado de máquina para prever e/ou caracterizar os doadores de sangue de modo que seja possível obter uma melhor compreensão sobre os fatores que influenciam o comportamento futuro do doador para poder prever esse comportamento com maior precisão?”.

Os artigos selecionados nesta RSL permitiram também responder as seguintes questões secundárias:

Q1 – Quais os algoritmos/técnicas mais utilizados nos trabalhos analisados?

Q2 – Quais características são mais relevantes para caracterizar o comportamento do doador?

Com o objetivo de responder os questionamentos foram elaboradas as *strings* de busca para realizar as pesquisas nas bases *online*. Durante a busca foram considerados os seguintes campos: título, palavras-chave e resumo. As palavras chaves da *string* de busca foram:

- *PREDICTION*
- *BLOOD DONOR*
- *BLOOD DONATION*
- *MACHINE LEARNING*

A *string* de busca utilizada foi:

(PREDICTION OR “MACHINE LEARNING”) AND (“BLOOD DONOR” OR “BLOOD DONATION”).

A Tabela 1 apresenta os critérios de inclusão e exclusão estabelecidos para a revisão. A inclusão de artigos levou em consideração o foco nas questões de pesquisa apresentadas.

² <https://pubmed.ncbi.nlm.nih.gov/>

³ <https://dl.acm.org/>

⁴ <https://ieeexplore.ieee.org/Xplore/home.jsp>

⁵ <https://link.springer.com/search>

⁶ <https://scholar.google.com.br>

Tabela 1: Critérios de seleção de trabalhos relevantes.

Critério	Identificador	Descrição
Inclusão	I1	Abordem a utilização técnicas de predição e/ou aprendizado de máquina para identificação de doadores de sangue que possam ter uma melhor taxa de fidelização.
	I2	Que foram publicados a partir de 2010 nas bases elencadas anteriormente.
Exclusão	E1	Trabalhos que não tratem exclusivamente de algoritmos/técnicas para a predição e/ou classificação de doadores de sangue.
	E2	Trabalhos que não foram publicados em língua inglesa ou portuguesa.
	E3	Artigos duplicados
	E4	Artigos com menos de 4 páginas.
	E5	Teses, dissertações e capítulos de livros.

3. Execução da Revisão

Aqui vale ressaltar que a busca limitou-se a no máximo os 50 primeiros artigos em cada base, pois, notou-se que após os 50 primeiros, os artigos apresentavam uma perda significativa de relevância. A Tabela 2 apresenta a quantidade de trabalhos recuperados, excluídos, escolhidos e lidos na íntegra utilizando os critérios mencionados anteriormente (Tabela 1) e a Figura 1 apresenta essa informação graficamente.

Tabela 2: Trabalhos escolhidos e suas bases de pesquisa.

Base de Busca	Artigos Recuperados	Artigos Excluídos	Artigos Escolhidos	Lidos na Íntegra
ACM	16	14	2	2
IEEE	5	4	1	1
Google scholar	50	40	10	10
SPRINGER	50	48	2	2
PUBMED	50	48	2	2

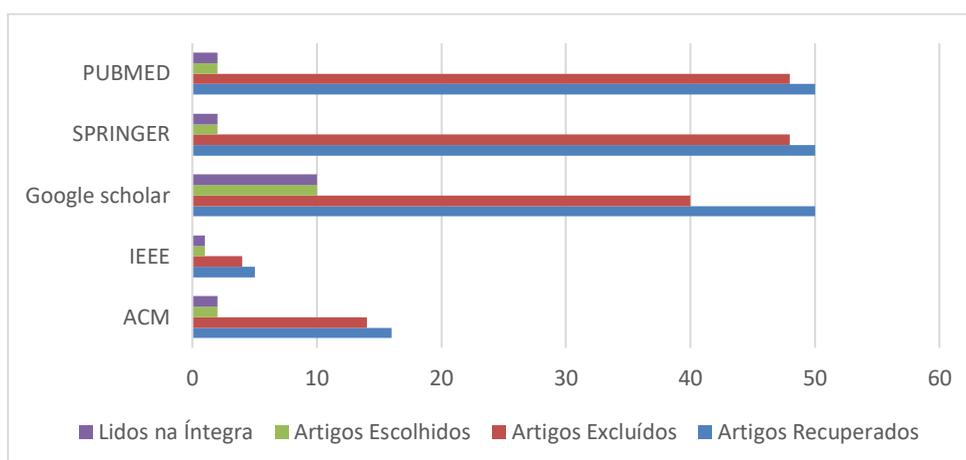


Figura 1: trabalhos recuperados por base

4. Resultados da revisão

Na Tabela 3, são apresentados os 10 trabalhos mais relevantes (com maior aderência) para a pesquisa, em ordem decrescente de ano de publicação, uma descrição dos mesmos e os algoritmos utilizados por cada um deles.

Tabela 3: Trabalhos mais relevantes

Referência	Título	Descrição	Algoritmos
[Concha, C. S.; Correa, P. M.] – 2021 – Google Scholar	<i>Predicting the Intention to Donate Blood among Blood Donors Using a Decision Tree Algorithm</i>	Tem como objetivo propor uma forma mais simplificada de mensurar um fenômeno social multideterminado, como a intenção de doar sangue novamente e a aplicação de árvores de decisão para alcançar essa simplificação, contribuindo assim para o campo da ciência de dados.	C4.5 (<i>Decision Tree</i>)
[Pabreja, K., Bhasin, A.] – 2021 – Google Scholar	<i>A Predictive Analytics Framework for Blood Donor Classification</i>	Procura identificar as características, em ordem de importância, que afetam a decisão de uma pessoa de se tornar um doador de sangue. As características utilizadas foram grupo sanguíneo, religião e 17 perguntas no padrão <i>likert</i> referentes a crenças do doador com relação ao ato de doar sangue.	K-NN (<i>K-Nearest Neighbor</i>), LR (<i>Logistic Regression</i>)
[Birhane, T.; Hailu, B.] – 2021 – Google Scholar	<i>Predicting the Behavior of Blood Donors in National Blood Bank of</i>	Aplicação de técnicas de mineração de dados em centros de saúde usando diferentes modelos. O processo de desenvolvimento	Naïve Bayes, J48 (<i>Decision Tree</i>), ANN (<i>Artificial</i>)

	<i>Ethiopia Using Data Mining Techniques</i>	do modelo ajuda a identificar ou prever o comportamento dos doadores de sangue, sejam eles elegíveis ou inelegíveis para doar sangue.	<i>Neural Network</i>)
[Alajrami, E.; Abu-Nasser, B. S.; Khalil, A. J.; Musleh, M. M.; Barhoom, A. M; Abu-Naser, S. S.] – 2019 – Google Scholar	<i>Blood Donation Prediction using Artificial Neural Network</i>	O objetivo desta pesquisa é estudar o desempenho do ambiente JustNN para cuidar da previsão de doação de sangue utilizando um modelo de rede neural artificial.	ANN
[Agarwal, K.; Gupta, M.; Gupta, K.; Khan, A; Nallakaruppan, M. K.] – 2019 – IEEE	<i>Blood Transfusion System Using Data Mining Techniques and GRA</i>	Apresenta a comparação de vários algoritmos, como árvore de decisão, floresta aleatória e regressão logística entre outros com a finalidade de descobrir o algoritmo com maior eficiência para ser utilizado no desenvolvimento de um sistema para prever a natureza de um doador de sangue, com base em dados anteriores de doação.	J48, CART (<i>Classification and Regression Trees</i>), RF (<i>Random Forest</i>), LR
[Shashikala, B. M.; Pushpalatha, M. P.; Vijaya, B.] – 2019 – Springer	<i>Machine Learning Approaches for Potential Blood Donors Prediction</i>	Utiliza algoritmos de aprendizado de máquina para prever se um indivíduo é um potencial doador a partir dos dados fornecidos pela pessoa tais como idade, sexo, grupo sanguíneo e um questionário com questões psicométricas.	K-NN, Naïve Bayes
[Alkahtani, A.; Jilani, M.] – 2019 – Google Scholar	<i>Predicting Return Donor and Analyzing Blood Donation Time Series using Data Mining Techniques</i>	Algoritmos de aprendizado de máquina, como regressão logística (LG), floresta aleatória (RF) e classificador de vetor de suporte (SVC) foram aplicados para desenvolver e avaliar modelos para classificar os doadores de sangue como doadores de retorno ou não.	LR, RF
[Bahel, D.; Ghosh, P.; Sarkar A.; Lanham, M.A.] – 2017 – Google Scholar	<i>Predicting Blood Donations Using Machine</i>	Estudo sobre o desempenho de algoritmos de aprendizado de máquina para dar suporte ao problema de previsão de doação de sangue.	CART, C4.5, ANN,

	<i>Learning Techniques</i>		SVM (<i>Support Vector Machine</i>), LR, LDA (<i>Linear Discriminant Analysis</i>)
[Wevers, A.; Wigboldus, D. H. J.; Kort, W. L. A. M.; Baaren, R.; Veldhuizen, I. J. T.] – 2014 – Google Scholar	<i>Characteristics of donors who do or do not return to give blood and barriers to their return</i>	Este estudo teve como objetivo investigar as características do comportamento de retorno do doador e obter informações sobre as barreiras à doação de sangue relatadas pelos próprios doadores. As características utilizadas foram idade, sexo e motivos para a não doação.	LR
[Santhanam, T.; Sundaram, S.] – 2010 – Google Scholar	<i>Application of CART Algorithm in Blood Donors Classification</i>	Procura identificar o comportamento que caracteriza a doação de sangue usando os algoritmos de classificação de mineração de dados. A análise foi realizada usando um conjunto de dados de transfusão de sangue padrão e usando um algoritmo de árvore de decisão.	CART

Na Figura 2, é apresentado um mapeamento dos algoritmos utilizados nos 10 trabalhos mais relevantes para a pesquisa. A regressão logística foi utilizada em 50% dos trabalhos, redes neurais artificiais e algoritmos de classificação e árvore de regressão em 30%, árvore de decisão C4.5, árvore de decisão J48, K-NN (K-Nearest Neighbor), Naïve Bayes e RF(*Random Forest*) em 20% e finalmente LDA (*Linear Discriminant Analysis*), SVM (*Support Vector Machine*) e árvore de decisão C5.0 em 10% dos trabalhos.

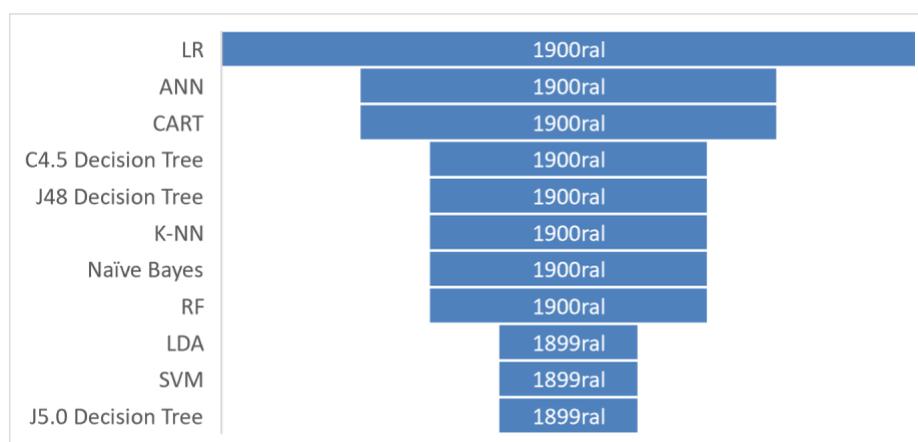


Figura 2: Mapeamento dos algoritmos utilizados nos trabalhos mais relevantes

Dentre os trabalhos mais relevantes destacam-se os seguintes:

- “*A Predictive Analytics Framework for Blood Donor Classification*” [Pabreja e Bhasin 2021] é um estudo sobre um grupo de cerca de 500 estudantes de uma Universidade do estado de Delhi, na Índia que responderam 20 questões (17 no padrão *Likert*) relacionadas ao desejo de doar de sangue. O estudo relata o uso de dois algoritmos de aprendizado de máquina populares: K-NN e LR (*Logistic Regression*). No trabalho apresentado nesse artigo é utilizada a ferramenta *Spyder*, um ambiente multiplataforma de software livre baseado em *Python* com o foco principal em análise de dados, sendo que o K-NN apresentou resultados mais interessantes com valores de 0,7037 e 0,7209 para acurácia e precisão respectivamente, contra 0,6815 e 0,0711 da LR.
- “*Application of CART Algorithm in Blood Donors Classification*” [Santhanam e Sundaram 2010] aborda o uso de árvores de classificação e regressão. O conjunto de dados utilizado foi uma base de dados com 748 doadores do Centro de Transfusão de Sangue de Hsin-Chu da Tailândia. As características dos dados presentes nessa base são: recência (meses desde a última doação), frequência (número total de doações), quantidade (total de sangue doado em ml), tempo (em meses desde a primeira doação) e uma característica binária que representa se ele / ela doou sangue em março de 2007 (campanha realizada pelo Centro de Transfusão). O modelo derivado do CART (*Classification and Regression Trees*), juntamente com a definição estendida para identificar doadores voluntários regulares, forneceu uma boa precisão de classificação neste conjunto de dados.

As características consideradas relevantes e utilizadas pela maioria dos algoritmos é a recência (meses que se passaram desde a última doação), o grupo sanguíneo, a frequência (número total de doações), a quantidade total de sangue doado (em ml.) e o tempo transcorrido desde a primeira doação. Outros algoritmos utilizaram as respostas de questionários no padrão *LIKERT* referentes a religião, tempo de duração da doação, sentimento do doador referente à possibilidade de contágio de uma doença ao realizar uma doação, recompensa além da satisfação pessoal etc. Um fato notado é que nenhum dos trabalhos tratou a distância entre o endereço do doador e o endereço do local de coleta

como uma característica com potencial para influenciar o comportamento dos doadores para se tornarem RVD, ficando assim como algo a ser pesquisado futuramente.

5. Conclusão

A revisão da literatura apresentada nesse artigo mostra que o uso de uma abordagem de aprendizado de máquina permite explorar quais características podem ser usadas para gerar uma previsão mais apurada sobre o comportamento futuro do doador de sangue e, portanto, possivelmente adicionar novas intuições sobre o que influencia o seu retorno. Alguns dos artigos recuperados na revisão identificam diversas características relevantes para incentivar doadores a se transformarem em doadores ativos.

As características utilizadas com maior frequência nas pesquisas analisadas são a idade, o sexo, a recência (meses desde a última doação), frequência (número total de doações), quantidade (total de sangue doado em ml) e tempo (em meses desde a primeira doação). Características culturais que podem influenciar na doação não são consideradas nos artigos recuperados.

Uma visão mais abrangente da aplicabilidade de técnicas apresentadas na seção 4, fundamentalmente da área de aprendizado de máquina, em geral, e mineração de dados, em particular, podem guiar também as políticas e decisões a serem adotadas para convocar doadores em épocas de crise ou em situações de emergência.

6. Referências

- Agarwal, K., Gupta, M., Gupta, K., Khan, A., Nallakaruppan, M. K. (2019). Blood Transfusion System Using Data Mining Techniques and GRA. In *2019 5th International Conference on Advanced Computing and Communication Systems (ICACCS)* (pp. 1143-1147). IEEE.
- Alajrami, E., Abu-Nasser, B. S., Khalil, A. J., Musleh, M. M., Barhoom, A. M., Naser, S. A. (2019). Blood donation prediction using artificial neural network. *International Journal of Academic Engineering Research (IJAER)*, 3(10).
- Alkahtani, S. A., Jilani, M. (2019). Predicting return donor and analyzing blood donation time series using data mining techniques. *International Journal of Advanced Computer Science and Applications*, 10(8).
- Bahel, D., Ghosh, P., Sarkar, A., Lanham, M. A. (2017). Predicting blood donations using machine learning techniques. *Computer Science*, 323.
- Birhane, T., Hailu, B. (2021). Predicting the Behavior of Blood Donors in National Blood Bank of Ethiopia Using Data Mining Techniques. *International Journal of Information Engineering and Electronic Business*, 13(3).
- Gillespie, T. W., Hillyer, C. D. (2002), Donation Decision. *Transfus. Med.*, vol. 16, no. 2, p. 115–130.
- Kitchenham, B. A., Brereton, P., Budgen, D., Turner, M., Khalil, M. (2007). Lessons from applying the systematic literature review process within the software engineering domain. *Journal of Systems and Software*, v. 80, n. 4, p. 571–583.

- Magalhães, T.A. et al. (2016), Prevalência de inaptidão sorológica dos doadores de sangue no hemocentro regional de Montes Claros, Minas Gerais. *Revista Online de Pesquisa*, vol. 8, no. 3, p. 4864–4871.
- Ministério da Saúde, Dezesseis a cada mil brasileiros doam sangue. 2019. <https://www.saude.gov.br/noticias/agencia-saude/45520-dezesseis-a-cada-mil-brasileiros-fazem-doacao-de-sangue>.
- Pabreja, K., and Bhasin, A. (2021). A Predictive Analytics Framework for Blood Donor Classification. *International Journal of Big Data and Analytics in Healthcare (IJBDAAH)*, 6(2), 1-14.
- Roberts, N., James, S., Delaney, M., Fitzmaurice, C. (2019). The global need and availability of blood products: a modelling study. *Lancet Haematol.*, vol. 6, no. 12, p. e606–e615, doi: 10.1016/S2352-3026(19)30200-5.
- Salazar-Concha, C., Ramírez-Correa, P. (2021). Predicting the Intention to Donate Blood among Blood Donors Using a Decision Tree Algorithm. *Symmetry*, 13(8), 1460.
- Santhanam, T., Sundaram, Shyam (2010). Application of CART algorithm in blood donors classification. *Journal of computer Science*, v. 6, n. 5, p. 548.
- Shashikala, B. M., Pushpalatha, M. P., Vijaya, B. (2019). Machine learning approaches for potential blood donors prediction. In *Emerging Research in Electronics, Computer Science and Technology* (pp. 483-491). Springer, Singapore.
- Wevers, A., Wigboldus, D. H., De Kort, W. L., Van Baaren, R., Veldhuizen, I. J. (2014). Characteristics of donors who do or do not return to give blood and barriers to their return. *Blood Transfusion*, 12(Suppl 1), s37.
- World Health Organization, Global Status Report on Blood Safety 2016. Geneva: WHO; 2017. <https://www.who.int/>.