[1.]Fatma Betül ARI, [1.]Haticenur BAĞDAT, [1.]Umut ÖZKAYA

BLOOD DONOR CLASSIFICATION BY USING ANN, KNN AND SVM

[1.] Konya Technical University Department of Electrical and Electronics Engineering, Konya, 42250, TURKEY

Abstract: In this study, Artificial Neural Networks (ANN), Support Vector Machine (SVM), K-Nearest Neighbour (KNN) algorithms were used to predict blood analysis. Comparative analysis was made with Machine Learning Methods. Accuracy, sensitivity, specificity, precision and F1 score were used in the study. The proposed ANN structure reached 79.86% accuracy, 91.74% sensitivity, 42.86% specificity, 83.33% precision and 87.34% F1 Score values in predicting blood analysis. This ANN structure showed the highest performance in predicting blood analysis, followed by KNN and SVM, respectively.
Keywords: Blood Donation, Artificial Neural Network, K-Nearest Neighborhood, Support Vector Machine

1. INTRODUCTION

In every day, people need blood transfusions as a result of diseases, injuries, accidents. Regular blood donating saves people's lives and also gives people hope for life. Another issue in blood donation is the accurate estimation of blood demand. If there is more blood than demanded, this excess blood will eventually complete its useful life and become unusable. Otherwise, loss of life may occur. As a result, the important thing is to maintain adequate blood supply. By using machine learning, accurate blood donation estimation can be made. Thus, it can be predicted whether a volunteer will be able to donate blood. In the literature, there have been more various approaches to solve this problem. Mohamad Darwiche et al. tried to improve blood donation prediction with machine learning algorithms. They used Multi-Layered Perceptron (MLP) and Support Vector Machine (SVM) algorithms. They conducted evaluations in 600 patients and prospectively confirmed in 148 patients. They achieved a sensitivity of 65.8% and a specificity of 78.2% in the prospective group [1]. Samy S. Abu-Naser et al. examined the performance of the JustNN over the blood donation problem. They created an Artificial Neural Network (ANN) model to measure performance. It was found that their JustNN model provides better performance accuracy (99.31%) than other studies [2]. Han Shih et al. predicted the blood supply efficiently in blood centers. Time series processing techniques and machine learning algorithms were used. Autoregressive (AUTOREG), Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA), Seasonal ARIMA, Seasonal Exponential Correction Method (ESM) and Holt-Winters models were preferred. ANN and multiple regression were tried in concept of machine learning algorithms. The lowest error measurements were observed in ESM and ARIMA models [3]. Shima Haghani et al. compared the accuracy of the proposed model in predicting the number of turns to blood donations with classical statistical models. 864 donors, who made a successful first donation, were followed for five years. The number of returns for blood donation was considered as the response variable. Poisson Regression (PR), Negative Binomial Regression (NBR), Zero Inflated Poisson Regression (ZIPR) and Zero Inflated Negative Binomial Regression (ZINBR) and ANN model were adapted to the data. Mean Squared Error (MSE) metric was used to compare models. STATISTICA 10 and R 3.2.2 were used to adapt the models [4]. Rohaifa Khaldi et al. tried to estimate three blood components monthly by using ANN. The predicted blood components were identified as red blood cells (RBC), plasma (CP), and platelets (PFC) [5]. B. M. Shashikala et al. estimated whether the person is a donor or not from the information provided by the people. Naive Bayes technique and K-Nearest Neighbor (KNN) algorithm are used. It has been observed that KNN accuracy value is higher than Naive Bayes algorithm [6]. Arunkumar Chinnaswamy et al. tried to distinguish those who donated blood from those who did not. They preferred Linear Discrimination Analysis (LDA), MLP and different machine learning applications. As a result, it has been determined that Probabilistic Neural Network (PNN) shows the highest accuracy of 100% [7]. Sadia Nadira Diba et al. designed a blood donation application in their study. In practice, they used a real-time map and machine learning algorithms [8]. WB Zulfikar et al. selected some criterias (blood type, gender, age) that a person who wants to donate blood must fulfill in their study. Decision tree and pure bayes classifications were used to determine the suitability of individuals. The accuracy value of 66.65% of the decision tree and 79.95% of the naive bayes were observed [9]. In this study, we proposed a model to predict blood donation. The data set is ready-used [10]. There are 576 samples in the data set. KNN, ANN and SVM algorithms are used. High performance (accuracy) has been

achieved with ANN. Section 1 is defined as introduction. Section 2 includes material and method. Results and Findings are belonged to Sections 3. Finally, Section 4 contains conclusion.

2. MATERIAL AND METHOD

— Dataset Description

The data set used in our study was obtained from the mobile blood donation device in Taiwan. According to the information shared by the Turkish Red Crescent, there are necessary conditions to donate blood. For example, the person who wants to donate blood, must be between the ages of 18-65. Women can donate in every four months; on the contrary, men donate in every three months. In whole blood donation, an average of 450 ml (± 10%) of blood is collected. There are some criteria such as the months since the last donation, the number of donations, the total volume donated (c.c.) and months since the first donation for blood donor. Donation information in March 2007 were collected. The number of samples was obtained 576. The number of donations varies with 1-50, the total volume donated (c.c) 250-1250 and the months since the first donation varies in 2-98. It has the value 1 if he was able to donate. Otherwise, it could be not.

— Artificial Neural Network

Artificial Neural Networks (ANN) have been developed by taking the human brain capability [11]. It is the software that produces new data by imitating the features of the human brain such as learning and predicting. ANN emerged in 1943 by Warren McCulloch and Walter Pitts [12].

x_i data is entered into the neurons. Inputs multiplied by their weights are added together. Bias is added to the multiplied data. As a result of the transaction, an output is obtained. As the results are passed through activation function, the data output is created.

$$z = b_0 + \sum w_i x_i \tag{1}$$

$$h\theta(x) = \frac{1}{1 + e^{-w^T x}} \tag{2}$$

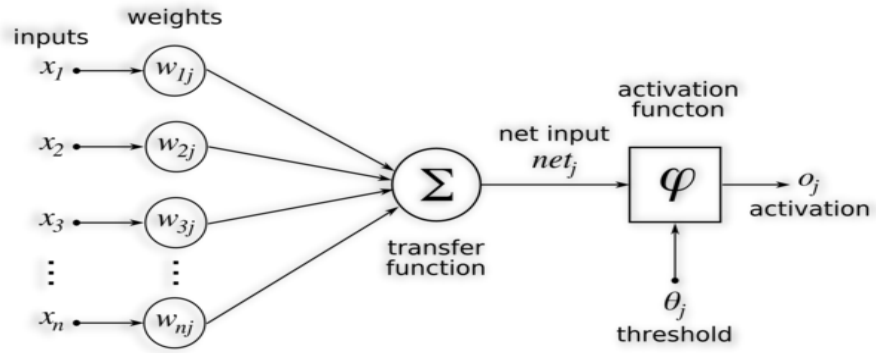

Figure 1. ANN Structure [13]

It was developed based on the insufficiency of single layer perceptron in nonlinear relationships. It consists of the input layer, hidden layers and the output layer.

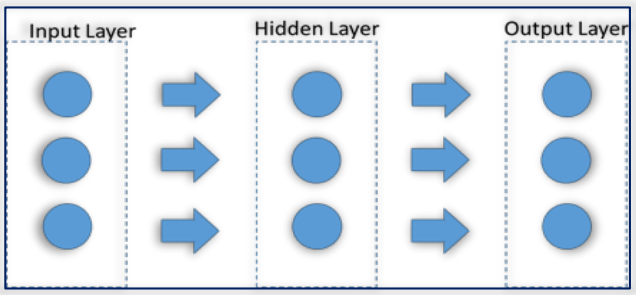

Figure 2. Multilayer Perceptron Model

— Support Vector Machine

Support Vector Machine (SVM) was proposed as a machine learning algorithm used for regression and classification challenges [15]. SVMs are often used in classification problems to divide into two classes. Linear decision boundaries are set to separate. It is intended that decision limit be at the maximum distance from any point in the training data.

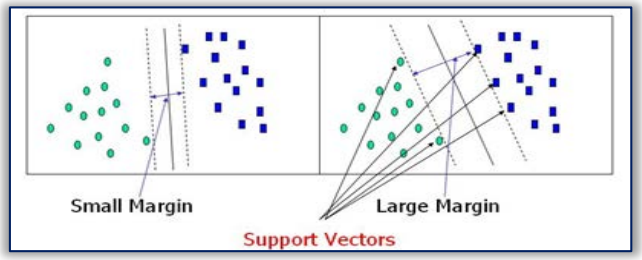

Figure 3. Support Vector Machine Decision Limit [16]

SVMs are divided into two types as linear and nonlinear SVM. It is the case of separating two classes distributed linearly with the help of the obtained decision function [17].

— **K-Nearest Neighbor Algorithm**

K-Nearest Neighbor Algorithm (KNN) algorithm is one of the sample in machine learning algorithms. It is classified according to the similarity between the samples in the training and testing dataset [18]. Neighborhood between similar data is high. Neighborhood decreases in non-similar data [19]. The most important parameters in the KNN algorithm are the number of neighbors (k), weighting and distance criteria. In the KNN algorithm, the number of neighbors as much as the k value is taken into account. For k = 1, it is assigned to its nearest neighbor.

3. RESULTS AND FINDINGS

576 samples in the data set were analyzed. The classification metrics were used in the comparative analysis such as accuracy, specificity, sensitivity, precision and F1-score. In addition, comparative analysis was carried out with different machine learning methods (ANN, KNN and SVM). All training and testing processes are prepared in Matlab 2020 a program. In the analysis, 75% of the classification data of blood donation is reserved for the training phase and 25% for the test phase. Five different metrics were used for blood donation classification in this study.

$$Accuracy = (TP + TN) / (TP + FN + TN + FP) \quad (3)$$

$$Specificity = TN / (TN + FP) \quad (4)$$

$$Sensitivity = TP / (TP + FN) \quad (5)$$

$$\Pr ecison = TP / (TP + FP) \quad (6)$$

$$F1\ Score = 2 \times TP / (2 \times TP + FN + FP) \quad (7)$$

TP, TN, FP and FN values are the number of correct positive, correct negative, false positive and false negatives, respectively. Table 1 includes results of different Number of Hidden Layer (NHL), Learning Rate (LR), Momentum constant (MC), Kernel Function (KF) and number of Neighborhood (N).

Table 1. Comparative Results

Method	Performance Metrics (%)				
	SEN	SPE	ACC	PRE	F1-Score
ANN (NHL:10 LR:0.9 MC:0.7)	91.74	14.29	72.92	76.92	83.68
ANN (NHL:15 LR:0.9 MC:0.7)	96.33	20	77.78	78.95	86.78
ANN (NHL:20 LR:0.9 MC:0.7)	95.41	17.14	20.28	78.2	85.95
ANN (NHL:20 LR:0.7 MC:0.7)	91.74	42.86	79.86	83.33	87.34
ANN (NHL:20 LR:0.5 MC:0.7)	98.17	8.57	76.39	76.78	86.29
ANN (NHL:20 LR:0.5 MC:0.5)	91.74	34.29	77.78	81.3	86.21
ANN (NHL:20 LR:0.9 MC:0.9)	96.33	11.43	75.69	77.21	85.71
SVM (KF:Gaussian)	99.09	2.94	76.39	76.76	86.51
SVM (KF:Linear)	99.09	0	75.69	76.22	86.17
SVM (KF:Polynomial)	100	2.94	77.08	76.92	86.96
KNN (N=1)	80.91	29.41	68.75	78.76	79.82
KNN (N=2)	90.91	17.65	73.61	78.13	84.03
KNN (N=3)	85.45	32.35	72.92	80.34	82.82

In this study, ANN KNN and SVM architectures were proposed. Machine learning methods were used to classify blood donation. As a result of the compared values, ANN provided the highest performance. With the proposed ANN structure, it reached 79.86% accuracy, 91.74% sensitivity, 42.86% specificity, 83.33% precision, 87.34% F1 score, 39.82% MCC values in predicting blood analysis. This ANN structure showed the highest performance in predicting blood analysis, followed by KNN and SVM respectively.

Note:

This paper is based on the paper presented at International Conference on Applied Sciences – ICAS 2021, organized by University Politehnica Timisoara – Faculty of Engineering Hunedoara (ROMANIA) and University of Banja Luka, Faculty of Mechanical Engineering Banja Luka (BOSNIA & HERZEGOVINA), in Hunedoara, ROMANIA, 12–14 May, 2021.

References

[1] M. Darwiche, M. Feuilloy, G. Bousaleh and D. Schang, "Prediction of blood transfusion donation," 2010 Fourth International Conference on Research Challenges in Information Science (RCIS), Nice, 2010, pp. 51-56

[2] E. Alajrami, A. Nasser, S. Bassem, J. Ahmed, M. Musleh, M. Barhoom, Alaa and S. S. Abu Naser, (2019). Blood Donation Prediction using Artificial Neural Network. International Journal of Academic Engineering Research (IJAER) 3 (10):1-7.

[3] H. Shih, and S. Rajendran, (2019). Comparison of Time Series Methods and Machine Learning Algorithms for

Forecasting Taiwan Blood Services Foundation's Blood Supply. Journal of Healthcare Engineering, 2019.
[4] Haghani S, Sedehi M, Kheiri S. Artificial Neural Network to Modeling Zero-inflated Count Data: Application to Predicting Number of Return to Blood Donation. J Res Health Sci. 2017;17(3):392. Published 2017 Sep 2.
[5] Rohaifa Khaldi, Abdellatif El Afia, Raddouane Chiheb, and Rdouan Faizi. 2017. Artificial Neural Network Based Approach for Blood Demand Forecasting: Fez Transfusion Blood Center Case Study. In Proceedings of the 2nd international Conference on Big Data, Cloud and Applications (BDCA'17). Association for Computing Machinery, New York, NY, USA, Article 59, 1–6
[6] Shashikala B.M., Pushpalatha M.P., Vijaya B. (2019) Machine Learning Approaches for Potential Blood Donors Prediction. In: Sridhar V., Padma M., Rao K. (eds) Emerging Research in Electronics, Computer Science and Technology. Lecture Notes in Electrical Engineering, vol 545. Springer, Singapore
[7] Chinnnaswamy, Arunkumar & V, Kamal & G, Gurusankar & Natarajan, Shabala. (2015). A study on automation of blood donor classification and notification techniques. international journal of applied engineering research. 10. 18503-18514.
[8] Diba, S.N. (2018). Blood donation application with implementation of machine learning.
[9] W. B. Zulfikar, Y. A. Gerhana and A. F. Rahmania, "An Approach to Classify Eligibility Blood Donors Using Decision Tree and Naive Bayes Classifier," 2018 6th International Conference on Cyber and IT Service Management (CITSM), Parapat, Indonesia, 2018, pp. 1-5
[10] [10]https://www.kaggle.com/bonastreyair/predicting-blood-analysis?select=blood-test.csv
[11] Ergezer, Halit, Mehmet Dikmen, and Erkan Özdemir. "Yapay sinir ağları ve tanıma sistemleri." PiVOLKA 2, no. 6 (2003): 14-17.
[12] Agatonovic-Kustrin, S., & Beresford, R. (2000). Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. Journal of pharmaceutical and biomedical analysis, 22(5), 717-727.
[13] Yılmaz, I., & Yuksek, A. G. (2008). An example of artificial neural network (ANN) application for indirect estimation of rock parameters. Rock Mechanics and Rock Engineering, 41(5), 781-795.
[14] Desai, K. M., Survase, S. A., Saudagar, P. S., Lele, S. S., & Singhal, R. S. (2008). Comparison of artificial neural network (ANN) and response surface methodology (RSM) in fermentation media optimization: case study of fermentative production of scleroglucan. Biochemical Engineering Journal, 41(3), 266-273.
[15] Zou, J., Han, Y., & So, S. S. (2008). Overview of artificial neural networks. Artificial Neural Networks, 14-22.
[16] Reis, H. Ç., & YILANCI, G. Destek vektör makineleri ve NDVI kullanarak pamuk ekili alanların tespiti: Harran ovası örneği. Türkiye Uzaktan Algılama Dergisi, 2(1), 29-41.
[17] Alkhatib, K., Najadat, H., Hmeidi, I., & Shatnawi, M. K. A. (2013). Stock price prediction using k-nearest neighbor (kNN) algorithm. International Journal of Business, Humanities and Technology, 3(3), 32-44.
[18] Özkaya, Umut & Seyfi, Levent. (2021). Evaluation of Ground Penetrating Radar B Scan Images via Convolutional Neural Network with Low Parameters. 6. 84-92

ISSN 1584 – 2665 (printed version); ISSN 2601 – 2332 (online); ISSN-L 1584 – 2665
copyright © University POLITEHNICA Timisoara, Faculty of Engineering Hunedoara,
5, Revolutiei, 331128, Hunedoara, ROMANIA
http://annals.fih.upt.ro