International Journal on
"Technical and Physical Problems of Engineering"
(IJTPE)
Published by International Organization of IOTPE

ISSN 2077-3528
IJTPE Journal
www.iotpe.com
ijtpe@iotpe.com

June 2020 | Issue 43 | Volume 12 | Number 2 | Pages 66-70

INVESTIGATION ON DIFFERENT PATTERN CLASSIFICATION METHODS AND PROPOSING THE OPTIMUM METHOD WITH IMPLEMENTATION ON BLOOD TRANSFUSION DATASET

Z. Jafari [1] A. Mahdavi Yousefi [1] S. Rajebi [2]

1. Department of Biomedical Engineering, Tabriz Branch, Islamic Azad University, Tabriz, Iran
stu.z.jafari@iaut.ac.ir, stu.a.mahdavi.iaut.ac.ir
2. Department of Electrical Engineering, Seraj Higher Education Institute, Tabriz, Iran, s.rajebi@seraj.ac.ir

Abstract- Blood transfusion is one of vital issues in medicine which is prescribed by a doctor in special cases. Blood transfusion service includes: collecting and distributing blood and its products by blood donation centers throughout the country. This organization is the only center providing blood and blood products to hospitals and other health centers. In developed and developing countries, extreme part of donators are built up by volunteers who don't receive any vantage because of cooperation that they have done. With this in mind, encouraging blood donation is an important goal in many countries. But only a constant rest period for those who donated blood cannot be a good measure for a person's readiness to donate blood again. Rather, you should use blood test results should be used to test person's readiness for blood donation. In this paper, using a blood transfusion service center's dataset, and using a variety of methods include Bayesian, KNN (*K*-Nearest Neighbor), Perceptron and RBF (Radial Based Function) Neural Network, the readiness or lack of readiness of individuals for blood donation is diagnosed. This is done by classifying people in two classes. People who have donated blood and those who have not donated blood.
The placement of people in a group who did not bleed indicates their readiness to donate blood again.

Keywords: Blood Transfusion, Blood Donation, Classification, Bayesian, KNN (*K*-Nearest Neighbour), Perceptron, RBF (Radial Basis Function), Neural Network.

1. INTRODUCTION

Blood donation happens that a healthy person voluntarily donates a certain amount of his blood at a blood transfusion center. After the screening of donated blood, it is used for provide blood products or is given to the other people through the blood bank of the treatment centers. The donation and injection process to recipients is called blood transfusion. Blood transfusion is needed for children and adults who have certain problems and illnesses.

A person who has an accident and loses a lot of blood or someone who has been operated or someone with a specific illness, such as thalassemia or cancer, or a type of anemia known as sickle cell anemia, are needed to blood transfusion.

Along with considering all the benefits of blood donation, one must also consider the person's readiness to donate blood. Therefore, there is a need to create a category for volunteers to donate blood. People who are ready to donate blood and those who are not yet ready to donate blood again.

In this paper, attempted to observe the importance of classification about readiness of volunteers. For this, randomly seven hundred forty-eight samples of donors are chosen from donor database. For all of these samples, the following features are collected: *F* (Frequency - number of all donations), *R* (Regency - months after last donation), *T* (Time - months after first donation), *M* (Monetary - total number of blood donating in c.c.), documents of 748 donors which include their blood donation information, whether he/she donated blood in March 2007.

The dataset of this study is collected from Hsin-Chu City's Blood Transfusion Service Center, in Taiwan). Predefined Classes in this paper, intended under two names: blood donated class and not blood donated class. The placement in the not blood donated class means the return of the person conditions to normal, or, in other words, the readiness of the person to re-donate blood.

In the following sections, some of supervised machine learning algorithms such as Bayesian, K-Nearest Neighbor, Perceptron and RBF Neural Networks are used to create a good classification of blood transfusion data under two categories; donating blood or not donating blood.

2. METHODOLOGY

Such as mentioned, different and famous methods of classification will be examined and compared. These methods, in the order of effectiveness, include: Bayesian, KNN, and Perceptron and RBF neural networks are the first defined and then coded and evaluated.

2.1. Bayesian Classifier

The basic assumption in this classifier is the existence of Gaussian distribution for all-natural data. For this purpose, the Gaussian distribution function of each of the classes is determined according to the mean and variance values of the training data. Then, the probability of placing the evaluated data (test data) is determined and then assigned to a class that is more possibility to be.

In accordance with the above, the functional flowchart of this classifier is considered and implemented according to Figure 1.

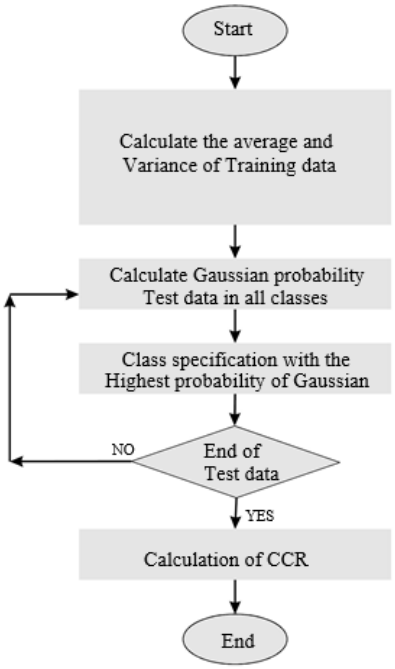

Figure 1. Flowchart of Bayesian classifier method

According to proposed, the Bayesian classifier has been programmed and after applying this classifier to the blood transfusion dataset, it was observed that the amount of *CCR* (Correct Classification Rate) is approximately *CCR*=53.5%.

The obtained *CCR* means that nearly one person is correctly classified from both. This is not promising and it's because of the lack of ideal Gaussian distribution for studied data. As a result, regardless of the Bayesian classifier, other classifications methods should be considered.

2.2. *K*-Nearest Neighbor Classifier

K-Nearest Neighbor classification method is one of the simplest classification methods. This method operates on voting and distance function. The metric used is the Euclidean distance. In this method, the class with the largest number of neighbors for test data, is selected its class. First, a certain amount is considered for *K*, and in terms of its value, *K*-Neighbours of the test data are considered. Which class has the largest neighbor among selected neighbours is reported as the class of the test data. The above algorithm is shown in the flowchart of Figure 2 in full.

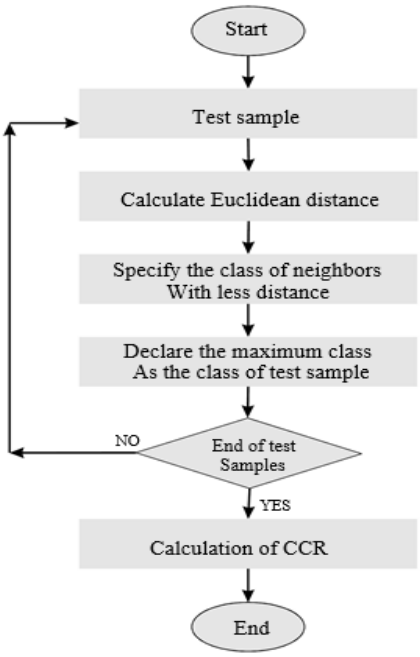

Figure 2. Flowchart of KNN classifier method

By choosing the most common value for *K* (*K*=3), and following the flowchart in Figure 2, the algorithm is implemented for our dataset and the *CCR* is reported as *CCR*=58.06%.

In this classifier, *K* is a determinant parameter. Common different values used for *K* in several papers (*K*=3, 5, 7) have been investigated and Correct classification rate for them are displayed in Figure 3.

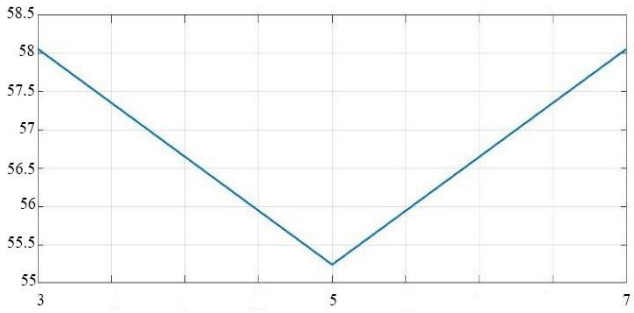

Figure 3. Effect of changing number of acceptable neighbors (*K*) in KNN method

The best of *CCR*s is happening when *K*=3 and *K*=7. But nevertheless, none of the results is acceptable, and thus this method, like the previous method, does not have a good ability to classify our data.

Considering unacceptable *CCR*s from Bayesian and KNN methods, neural networks will be proposed as new classification methods in the following sections to achieve better results.

2.3. Neural Network

Neural Network algorithm is a mathematical model including of some interconnected processing functions named neuron that categorized in input, output and hidden layers, which are like the structure of the human brain. The Artificial Neural Network may be considered

as an effective factor parallel distributed functions which has a natural tendency to store experimental knowledge and use it for the intended operation. One of the most important applications of the neural network is in the classification of statistical data.

In the following section two famous neural networks in classification (Multi-layer Perceptron and Radial Basis Function) are presented.

2.3.1 Perceptron Neural Network Classifier

The simplest model of neural networks is the Multi-Layer Perceptron (MLP) which simulates the transfusion function of the human brain. This algorithm is one of the first artificial neural networks that have been used for linear classifications. Perceptron is a machine learning algorithm that is placed in supervised learning category. The Perceptron algorithm is a type of classification that can decide, depending on the input vector, whether that entrance belongs to a class or not such as shown in Figure 4, MLP structure has an input layer, an output layer, and one or more hidden layer(s) between the input and the output layers. The layers are linearly interconnected with optimized values of coefficients called weight coefficients.

In each layer there are transformation functions with the name of the neuron whose output is under the transformation function of their input. The number and function of these neurons, together with the values of weights, are the basic parameters of the proper functioning of the neural network.

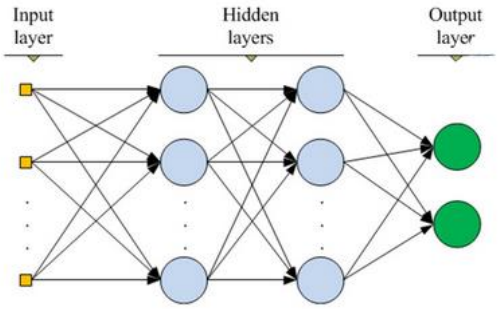

Figure 4. Multi-layer perceptron neural network

First, such as shown in Figure 5, a simple perceptron neural network with one hidden layer has been designed. For this network, number of neurons for hidden layer has been selected to 13 and (tan sig) and Purelin functions are considered as hidden and output layer's transfer functions respectively. Common used transfer functions in Multi-Layer Perceptron networks are (tan sig), Purelin and (log sig). Chart of these function is shown in Figure 6.

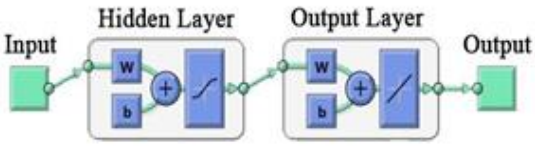

Figure 5. Proposed structure for one-layer perceptron neural network

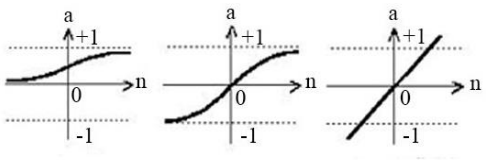

Figure 6. Various used transfer functions for perceptron NN

An overview of the perceptron neural network shown in Figure 4 was trained and then evaluated for our dataset classification. The result for this simple structure is reported as *CCR*=44.6%. The regression line for this classification is shown in Figure 7. The extreme deviation of regression line from the graph bisector indicates a lack of accuracy in the classification of this neural network.

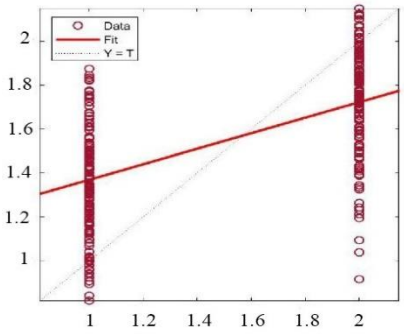

Figure 7. Regression line for simple one hidden layer perceptron

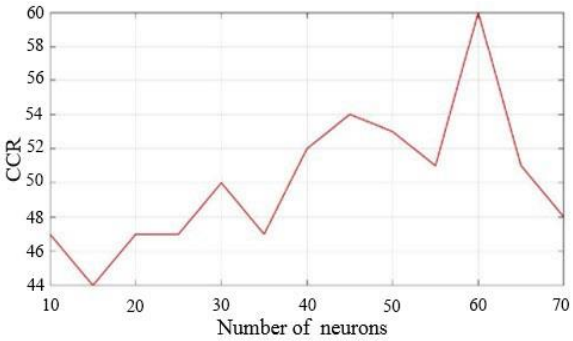

Figure 8. Effect of size of neurons in middle layer of Perceptron neural network on *CCR*

Amount of neurons in hidden layer of MLP neural networks has important effect on performance of the MLP network. Correct classification rate of the one hidden layer MLP, has been investigated for different values of size of the hidden layer. This assessment is shown in the curve of figure 8. According to this figure, the best *CCR* by using 60 neurons in hidden layer is equal to 60.52%

Using the second hidden layer and changing the transform functions of the neurons are known as solutions to improve performance of MLP network as a classifier. According to the above, a MLP neural network with two hidden layers was used such as depicted in Figure 9. different transform functions and the quantity of neurons for its middle (hidden) layers were considered and its classification accuracy was evaluated. The best obtained *CCR* for this network was equal to 60.38%.

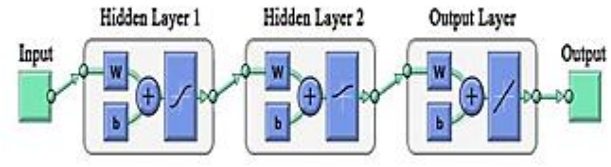

Figure 9. Proposed structure for two-layer perceptron neural network

Despite the relative improvement in the accuracy of classification in the MLP neural network compared to the Bayesian and KNN methods, the obtained *CCR* is not desirable. So as the last solution for successful classification of our dataset, RBF neural networks are proposed in the following.

2.3.2 Radial Basis Function Neural Network Classifier

RBF is another model of Artificial Neural Networks that utilize radial basis functions to approximate the separator classes of each other. In the other words, the output of this system is linear combination of radial basis functions and input vectors. In this neural network, network training means adjusting the mean and variance values of the Gaussian functions (type of RBFs) to cover the classification line.

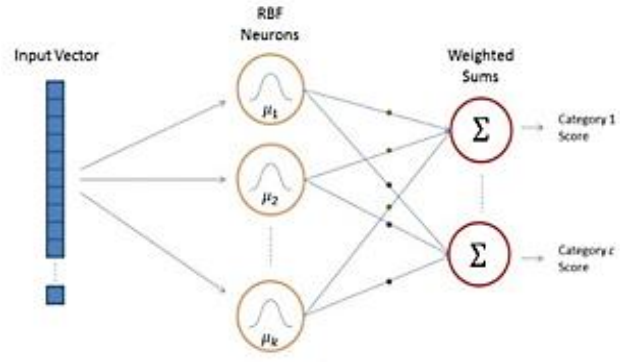

Figure 10. Radial based function neural network

Obviously, increasing the number of these basis functions makes the classification line more accurately modelled. This has the advantage of this neural network. In the Perceptron neural networks, increasing the number of neurons is not guaranteed to improve the performance of the classifier, but in the RBF neural networks, increasing the amount of radial basis functions will improve the performance of the network. Therefore, by choosing a sufficient amount of radial basis functions, the accuracy of the classroom can be increased.

The structure of a RBF Neural Network is shown in Figure 11. The used radial basis functions for this network have been chosen the Gaussian functions. The amount of used Gaussian functions (Radial Based Function) is selected to 500. The proposed RBF structure has been simulated and after training, its correct classification rate has been reported as 93.21%. Also, regression line of this network is shown in Figure 12.

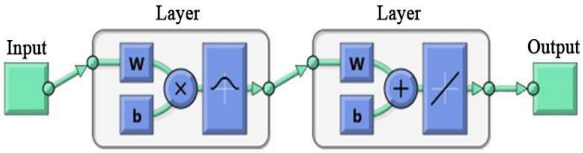

Figure 11. Proposed Structure for RBF neural network net

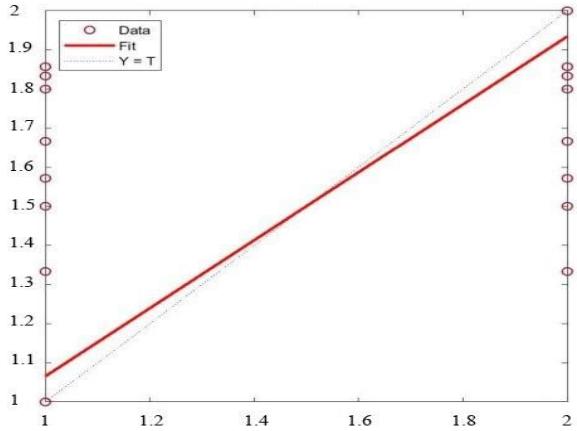

Figure 12. Regression line of proposed RBF neural network

Amount of Radial Based Function which will have remarkable effect on *CCR* classifier has been investigated and affected of it on *CCR* is reported as a Figure 13. As stated in Figure 12, we can consider 300 Gaussian functions as optimum number of the RBFs.

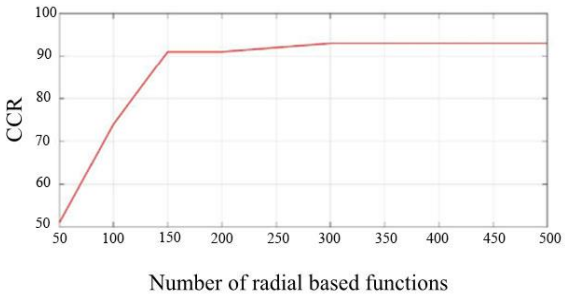

Figure 13. Effect of numbers of radial basis functions on *CCR* of proposed RBF Neural Network

3. CONCLUSION

The purpose of this paper was to find an effective and accurate classifier to identify people who are prepared to donate blood with people who are not prepared for it. Placement in the not blood donated group means the person's readiness for re-donating blood. Also, placing a person in a blood donation group means that there is a deficiency in the body of the person and, as a result, his lack of readiness for re-blooding.

First, Bayesian and *K*-NN methods were used for this classification. The acquired results were not considerable. The inaccuracy of these two methods can be due to two reasons: Not so much training data and the lack of proper distribution of Gaussian data. It should be noted that in the Bayesian method, the appropriate Gaussian distribution of data and in the neighboring method a large number of educational data plays a decisive role.

Then the neural networks were introduced and multi-layer Perceptron was used as classifier. In this method, after optimization on size, number of layers and transfer functions, highest rate of *CCR* was reported as 60%.

Finally, the RBF method was used as final classifier. Correct classification rate of this type of neural network was 93.21%. That was the best accuracy for the classification and thus the RBF Neural Network can be suggested as the best classifier for donating or not donating datasets.

REFERENCES

[1] N.H. Rajini, R. Bhavani, "Classification of MRI Brain Images Using K-Nearest Neighbor and Artificial Neural Network", IEEE, ICRTT, 2011.

[2] M. O'Farrell, E. Lewis, C. Flanagan, N. Jackman, "Comparison of K-NN and Neural Network Methods in the Classification of Spectral Data from an Optical Fibre-Based Sensor System Used for Quality Control in the Food Industry", Sens. Actuators B: Chemical, 111-112C, pp. 354-362, 2005.

[3] R.O. Duda, P.E. Hart, D.G. Stork, "Pattern Classification", Wiley, New York, 2001.

[4] M.A. Mazurowskia, P.A. Habasa, J.M. Zuradaa, J.Y. Lob, J.A. Bakerb, G.D. Tourassib, "Training Neural Network Classifiers for Medical Decision Making: The Effects of Imbalanced Datasets on Classification Performance", Neural Network, Vol. 21, No. 2-3, pp. 427-436, 2008.

[5] J. Tang, C. Deng, G. B. Huang, "Extreme Learning Machine for Multilayer Perceptron", IEEE, 2015.

[6] S. Haykin, "Neural Networks: A Comprehensive Foundation", Prentice Hall, 1999.

[7] C.K.I. Williams, D. Barber, "Bayesian Classification with Gaussian Process", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 20, Issue 12, Dec. 1998.

[8] I.Ch. Yeh, K.J. Yang, T.M. Ting, "Knowledge Discovery on RFM Model Using Bernoulli Sequence", Expert Systems with Applications, 2008.

[9] U. Orhan, M. Hekim, M. Ozer, "EEG Signals Classification Using the K-means Clustering and a Multilayer Perceptron Neural Network Model", Expert Systems with Applications, No. 38, pp. 13475-13481, 2011.

[10] S. Walczak, J.E. Scharf, "Reducing Surgical Patient Costs Through Use of an Artificial Neural Network to Predict Transfusion Requirements", Decision Support Systems, No. 30, pp. 125-138, 2000.

[11] D.F. Specht, "Probabilistic Neural Networks", Neural Networks, Vol. 3, pp. 109-118, 1990.

[12] B. Widrow, M.A. Lehr, "30 Years of Adaptive Neural Networks: Perceptron, Madaline, and Backpropagation", Proceedings of the IEEE, Vol. 78, No. 9, September 1990.

[13] R.P.L. Ippmann "Pattern Classification Using Neural Networks", IEEE Communications Magazine, Vol. 27, Issue 11, November 1989.

[14] D. Lowe, "Adaptive Radial Basis Function Non-Linearities and the Problem of Generalization", IEE Conf. on Artificial Intelligence and Neural Networks, 1989.

[15] L.J. Yan, N. Cercone, "Bayesian Network Modelling for Evolutionary Genetic Structures", Computers and Mathematics with Applications, No. 59, pp. 2541-2551, 2010.

BIOGRAPHIES

Zahra Jafari was born in Maragheh, Iran, 1994. She received the B.Sc. degree from Department of Biomedical Engineering, Tabriz Branch, Islamic Azad University, Tabriz, Iran in 2018. Currently, she is studying a M.Sc. at the same university. Her research interest are optimization, neural network and classification.

Asma Mahdavi Yousefi was born in Tabriz, Iran, on March 23, 1996. She received the B.Sc. degree from Department of Biomedical Engineering, Tabriz Branch, Islamic Azad university, Tabriz, Iran, in 2018. Currently, she is studying a M.Sc. at the same university. Her research interests are optimization, neural network and classification.

Saman Rajebi was born in Tabriz, Iran in 1981. He received his B.Sc. degree in Electrical Engineering in 2003 and M.Sc. and Ph.D. degrees in Communication Engineering in 2006 and 2017, respectively. He is a Fulltime Teaching Staff of Seraj Higher Education Institute, Tabriz, Iran and is the member of Executive Committee of ICTPE conferences. His research interests areas are radio frequency electronics, microwave applications, antenna designing and neural networks.