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

March 2020 | Issue 42 | Volume 12 | Number 1 | Pages 114-120

EVALUATION OF K-NEAREST NEIGHBOR, BAYESIAN, PERCEPTRON, RBF AND SVM NEURAL NETWORKS IN DIAGNOSIS OF DERMATOLOGY DISEASE

V. Yousefi [1] S. Kheiri [1] S. Rajebi [2]

1. Department of Biomedical Engineering, Tabriz Branch, Islamic Azad University, Tabriz, Iran
stu.s.kheiri@iaut.ac.ir, stu.v.yousefi@iaut.ac.ir
2. Department of Electrical Engineering, Seraj Higher Education Institute, Tabriz, Iran, s.rajebi@seraj.ac.ir

Abstract- In 1808, an English doctor named Robert Willan published a book on skin diseases in London, The field of specialized skin and modern dermatology. The advancement of technology and the emergence of laboratory methods and new research tools, the 21st century has become a century of skin disease treatment. In medical science, timely detection and diagnosis can prevent disease and save people. The purpose of data mining is clinical modeling which will help doctors improve their prevention, diagnosis and treatment programs and classification in the disease causes a more accurate and faster diagnosis of the disease. We want to use different classification methods for automatic detection and without the intervention of the disease class and in this paper, various methods of classification including method, KNN, Bayesian, neural network and the SVM used in this classification are relatively new method .In recent years, the support vector machine method has shown to perform better than more sophisticated methods for classification but the best answer to the obtained data was derived from the neural network method [1].

Keywords: Classification, KNN, Bayesian, Neural Network, SVM.

1. INTRODUCTION

Dermatology is a branch of medicine that works with the skin and its diseases. A dermatologist treats skin, scalp, nail and hair diseases. In this specialty, there are about 3,000 disorders and illnesses that encompass a wide range of cognitive impairments and some of them have a high prevalence rate in communities. The importance of these diseases is ignored in terms of public health and community health and for reasons of neglect, this leads to many adverse consequences for public health. The disease can have different types. Type of skin disease and dermatology is shown in Figure 1.

The characteristics of the disease are used as criteria for the classification of disease which is as follows:

For example, the medical history of a family can be of great help to family members in the field of medicine in the future because family members can cope with diseases that threaten them with information about their family's medical history and we can save people by identifying illnesses in a timely manner like heart disease, high Blood pressure, stroke, certain cancers and diabetes Age-related diseases include cardiovascular disease, cancer, cataracts, osteoporosis, type 2 diabetes, high blood pressure, and Alzheimer's disease.

Patients in this study were studied with characteristics and examples of these characteristics were studied [3].

Figure 1. Type of skin disease and dermatology [2]

2. METHODOLOGY

In this paper, we intend to classify 5 data classes that contain 34 features that are used for analysis and classify the disease by KNN, Bayesian, neural networks and SVM.

2.1. K-Nearest Neighbor Classifier

The KNN algorithm can be used for classification and regression issues and the KNN algorithm is one of the simplest classification algorithms that calculates Euclidean distance of test data from educational data. The test data class is selected based on this interval and so that the class that has the largest number of neighbors is determined as the test data class [4]. The KNN flowchart is presented in Figure 2. The KNN method is scheduled with the flowchart shown in Figure 3 and after simulation the *CCR* is equal to 14.66%.

The impact of neighboring numbers on *CCR* has been measured. In number of three neighbors: *CCR*=18.66%, in the number of five neighbors: *CCR*=14.66% and in the seven neighbors *CCR*=13.33%. It is obvious that the amount of *CCR* is low and it is not acceptable at all. The change of *K* doesn't affect considerably on the *CCR* and thus it can be said that this classifier does not have a good performance on recognition of class of dermatology dataset.

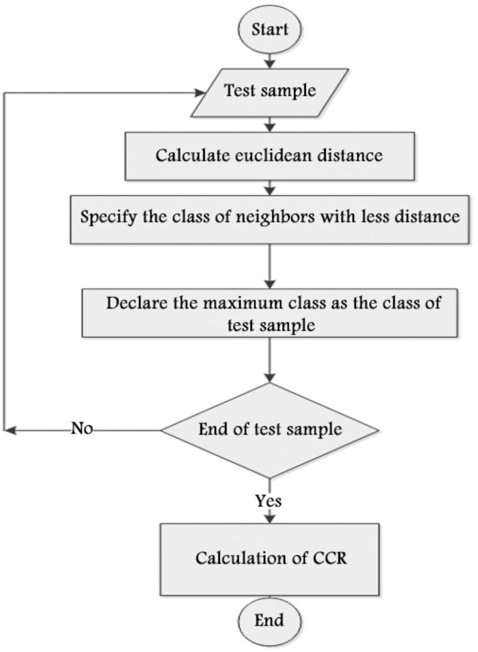


Figure 2. Flowchart for programming the KNN method

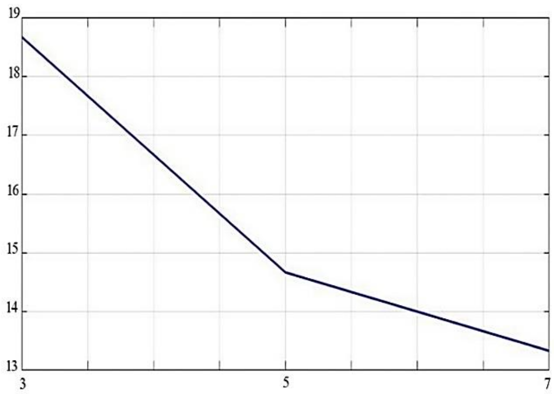


Figure 3. Amount of *CCR* of KNN classifier with changing in *K*

2.2. Bayesian Classifier

In the Bayesian method, considering the Gaussian distribution for all training data in each class, based on the Gaussian probability density distribution function such as given in Equation (1), the probability of the test data in each class is measured and data is assigned to a class that in which has the highest probability value.

A view of a Gaussian function with its parameters is shown in Figure 4.

As shown in Figure 4, the mean value and mean of the variance of the data are required to calculate the probabilities of the Gaussian function.

For this reason, the elementary part of this method is to calculate the average value and mean of the variance of educational data.

The functional flowchart of the Bayesian method is shown in Figure 5.

$$p = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(x-\mu)^2}{\sigma^2}} \tag{1}$$

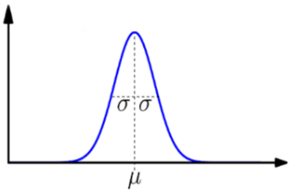


Figure 4. Gaussian distribution function

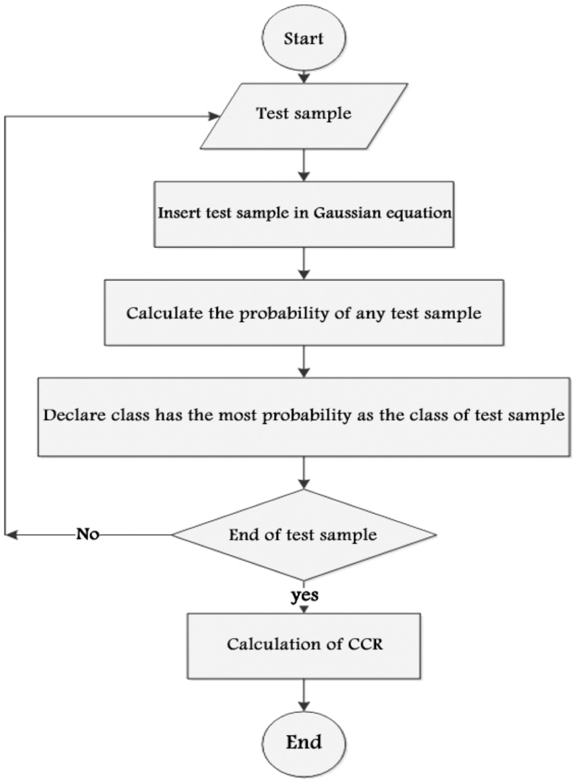


Figure 5. Flowchart for programming the Bayesian method

Based on the flowchart shown in Figure 5, the Bayesian classifier has been programmed and after simulation the CCR of this method is equal to 53.33%.

It can be seen that Bayesian's method has better CCR than the KNN method but its CCR value is not acceptable yet.

2.3. Neural Network Classifier

A neural network is inspired from human brain cells to enable learning and decision making. Like to human brain, artificial neural networks have a complex system that combines this complexity with multiple connections between its components. The neural network systems can solve many of problems without the need to know how to analyze that problem [5].

The only need for these structures for proper operation is the need for training for them. This is similar to the performance of the human brain's neural networks, which does many things without being aware of their relationship. The mathematical model for an artificial neural network is depicted in Figure 6.

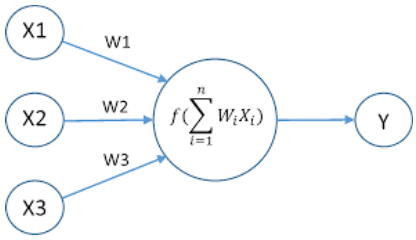


Figure 6. Mathematical structure of neural networks

2.3.1. Perceptron Neural Network

One of the most basic models available for neural networks is the Perceptron neural network. The input is a vector Perceptron of true values and in the hidden layer, a linear combination of these inputs is calculated which ultimately gives us the output. In the artificial neural network model, we assign a weight to each input that these weights are really important inputs for us. Adding bias also makes it easier to use Perceptron network. The output in this network is a combination of weights and according to the weight values, the output is obtained. Figure 7 shows a single-layer Perceptron neural network [6].

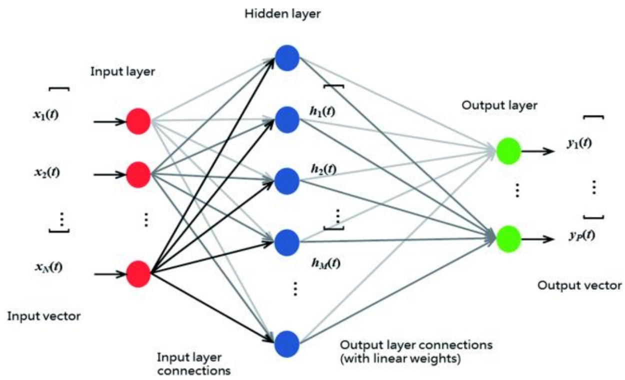


Figure 7. Structure of a single-layer Perceptron [7]

A single-layer neural network whit 25 neurons have been investigated and the network is shown in Figure 8. The number of neurons affects the Perceptron neural network and this effect is shown in Figure 9.

Transfer functions affect the performance of the network and some of commonly used transfer functions for neural networks are shown in Figure 10.

The result of this network operation as regression is shown in Figure 11 and at best CCR = 99.65%. The effect of the transformation functions whit 25 neurons on the network performance is investigated and shown in Table 1.

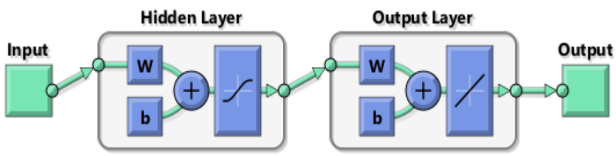


Figure 8. Structure of proposed one hidden layer Perceptron NN

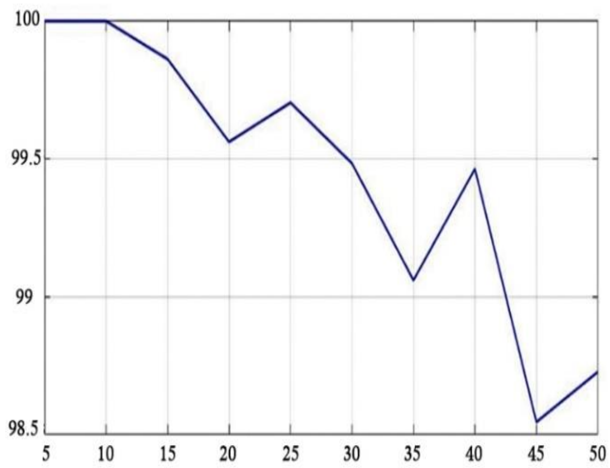
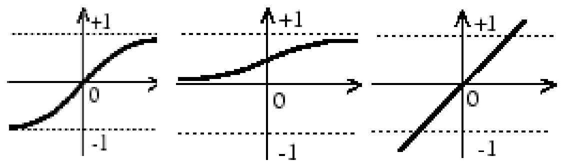


Figure 9. Effect of number of neurons in hidden layer of Perceptron



(a) (b) (c)

Figures 10. Commonly used transfer functions in neural networks (a) ansig, (b) logsig, (c) purelin

Table 1. The effect of three transformation functions on regression

Various conversion functions	tansig	logsig	purelin
Simulated CCR	%99658	%99847	%99959

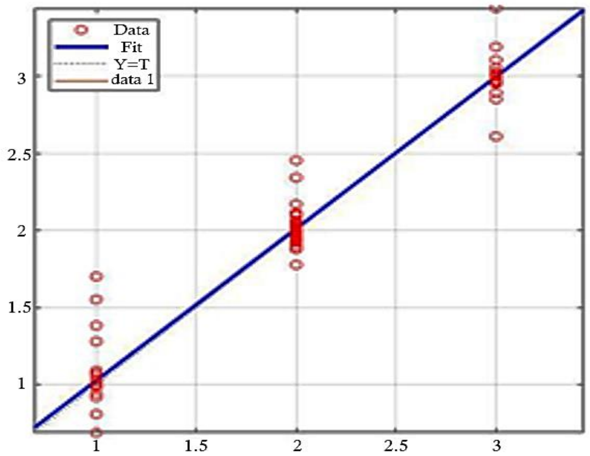


Figure 11. Regression achieved in the method Perceptron

2.3.2. Radial Basis Function of Neural Network

RBF neural networks are based on interpolation theories that consist of three leading layers. The first layer received an input vector and it propagates to the mid layer, which consists of neurons that use radial base functions and the output is received in the last layer. This network has a quick learning and preparation process and has the ability to approximate the various types of functions [8, 9]. The structure of this network is shown in Figure 12.

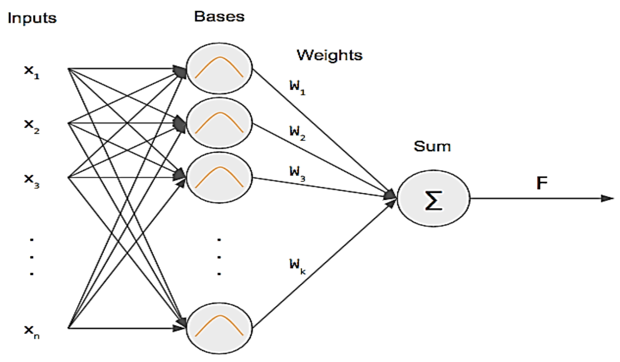


Figure 12. Commonly Structure of RBF Neural Network [10]

An RBF neural network has been investigated with 200 radial base functions and the network is shown in Figure 13. The number of radial base bases used in the RBF neural network affects the amount of CCR. This issue has been investigated and shown in Figure 14.

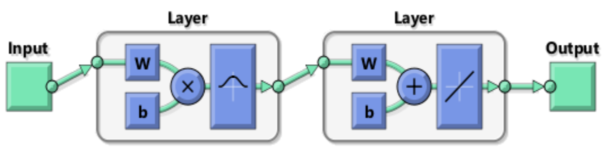


Figure 13. RBF neural network structure

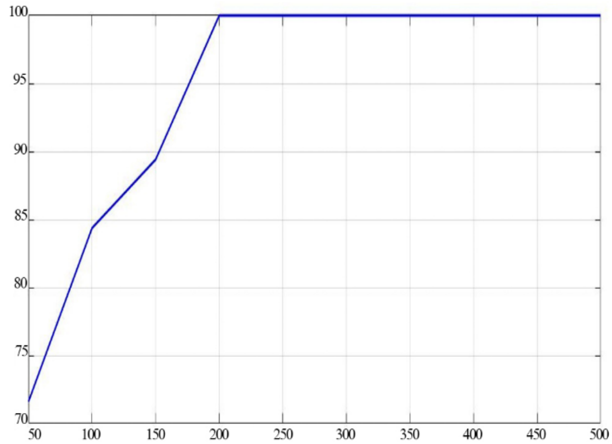


Figure 14. Effect of Radial Base Functions on RBF Neural Network

The number of different radial base functions that has been used indicates that 200 is the best answer. The simulated regression with 200 radial base functions is shown in Figure 15.

According to Figure 15, which shows the best regression of the network, CCR = 99.99%. Performance on the proposed RBF Neural Network is shown in Figure 16. This figure shows how the error rate of this classifier reduces by adding the Radial Basis Function to structure of the RBF Neural Network.

Execution time for different classification methods are shown in Table 2. (Processor: core i5CPU@1.70GHzRAM6.00)

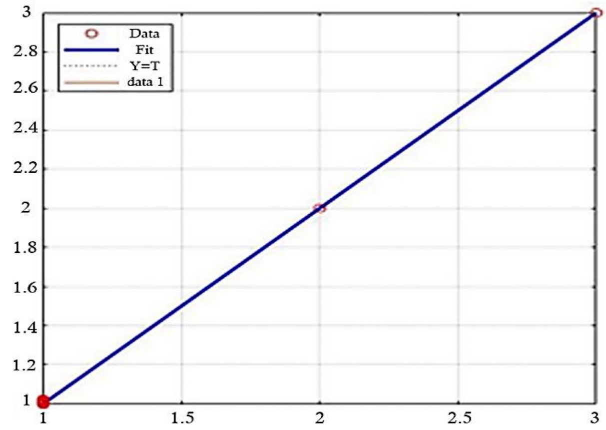


Figure 15. Regression achieved in the method RBF

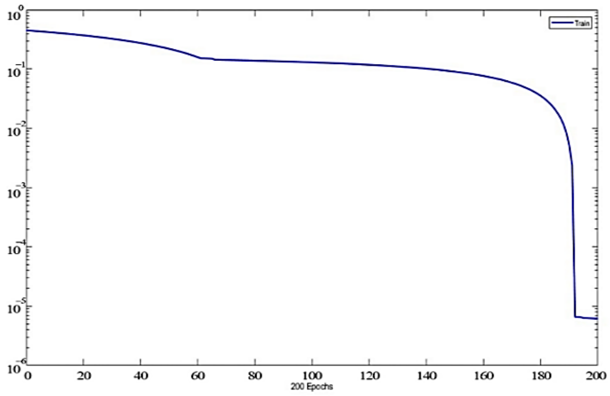


Figure 16. Performance of proposed RBF Neural Network

Table 2. Execution times for different classification methods

Classification methods	KNN	Bayesian	Perceptron	RBF
Execution Time	0.0717	0.0889	1.997	6.272

2.4. Support Vector Machine Classifier

SVM is a set of machine learning algorithms which is generalized to the nonlinear mode. It can be used for classification and regression in the dataset. This algorithm can be an effective method for data modeling by increasing the problem size and using the kernel function.

The purpose of this algorithm is to find the best boundary between the data So that it has the most distance from all categories. The SVM method usually delimits two classes but for multi-class datasets, the classes are compared in pairs to find the desired class.

If the categories are linearly separated, Obtains hyperplanes with maximum margin to separate categories. But data that is not linearly separable, the data is mapping to a larger dimension space to separate them in this new space linearly.

The separation of the two classes using the SVM classification method, optimal hyperplane and maximum margin is shown in Figure 17.

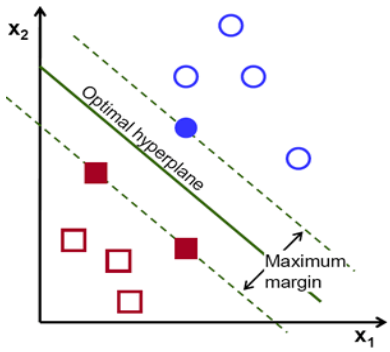


Figure17. Separation in two classes [11]

Objective is Find the best line (hyperplane) that separates the two categories. In the two-dimensional case the equation of this line is described by the Equation (2):

$$w_1X_1 + w_2X_2 + b = 0 \tag{2}$$

In the above equation w denotes the slope and b denotes the line spacing from the origin. The next *n* state is shown by relation (3):

$$w^T x + b = 0 \tag{3}$$

Equation (3) divides the page space into two parts, where the equations of the separated parts are shown by Equation (4):

$$\begin{aligned} &w^T x + b > 0 \\ &w^T x + b < 0 \end{aligned} \tag{4}$$

Instead of using this line in SVM, consider two parallel lines to create a more secure border with relation (5) the following:

$$\begin{aligned} &w^T x + b = 1 \\ &w^T x + b = -1 \end{aligned} \tag{5}$$

Suppose we have a set of *n* data:

$$\{x_i, y_i\};\ i = 1, 2, ..., n$$

$$x_i \in R,\ y_i \in \{-1, 1\}$$

If $y_i = 1 \Rightarrow w^T x_i + b > 1$

If $y_i = -1 \Rightarrow w^T x_i + b < -1$

Given the value of *y* and multiply it by the above equation, relation (6) is obtained:

$$y\left(w^T x + b\right) > 1 \tag{6}$$

The greater the distance between two parallel lines, the greater the margin of confidence and the better the separation works, we call the distance of two parallel lines from the center line *d*. The best state of to separate the classes is to have these two spaces equal $d_1 = d_2$.

This distance 'd' is calculated by Equation (7):

$$d_1 = \left| \frac{-(b-1)}{w} - \frac{-b}{w} \right| = \frac{1}{w} \tag{7}$$

Since $d_1 = d_2$ is state of best, Equation (8) is obtained.

$$d = d_1 + d_2 = \frac{2}{w} \tag{8}$$

For $d_1 = d_2$ maximum to be, w must be minimum or $d_1 = d_2$ must be a minimum.

The goal is to find $\min \frac{1}{2} w^T w$ with the $y\left(w^T x + b\right) > 1$ condition.

The solution is a Quad programming method. In this way, state of the equation must be saddle point:

$$L = \frac{1}{2} w^T w - \sum_i^n \alpha_i y_i \left(w^T x_i + b\right) - 1 \tag{9}$$

In the relation (9), L is the saddle equation and must be a minimum.

For *L* to be minimum, we must derive *L* from *w* and *b* and set it to zero. According to relations (10) and (11):

$$\frac{\partial L}{\partial w} = w - \sum \alpha_i y_i x_i = 0 \Rightarrow \quad w = \sum \alpha_i y_i x_i \tag{10}$$

$$\frac{\partial L}{\partial b} = \sum \alpha_i y_i = 0 \tag{11}$$

By assigning these values to the problem *L*, the equation changes according to relation (12):

$$L = -\frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j y_i y_j X_j^T X_i \tag{12}$$

must be α_i greater than zero so that *L* will eventually be minimized.

According to relation (13):

$$L = \sum_i \sum_j \alpha_i \alpha_j y_i y_j X_j^T X_i - \sum_i \alpha_i \tag{13}$$

By considering relations (14), (15), (16) and (17):

$$y_i y_j X_j^T X_i = h_{ij} \tag{14}$$

$$\frac{1}{2} \sum_i \sum_j \alpha_i \alpha_j h_{ij} = \alpha H \alpha^T \tag{15}$$

$$-\sum_i \alpha_i = F\alpha \tag{16}$$

$$L = \alpha H \alpha^T + F\alpha \tag{17}$$

where, *F* is a matrix with all its layers -1.

In MATLAB the quad program solves this problem and gives us the α answer which *w* and *b* can be calculated by the relations (18) and (19):

$$w = \sum_i \alpha_i y_i x_i \tag{18}$$

$$\begin{aligned} &y(w^T x + b) > 1 \rightarrow y^2 (w^T x + b) > y \\ &b_i = y_i - w^T x_i \Rightarrow b = \text{mean}(b_i) \end{aligned} \tag{19}$$

and the average of all of these b_i determines the bias coefficient *b*. The obtained *w* and *b* specify the boundary.

What has been said so far is about hard margin SVM and in general it is not applicable because a classification is purely linear. This method works 100%, which is not always good.

To make the method workable, we introduce the soft margin SVM:

Soft margin allows the program to have slight error and the separator boundary in this method is a curve as shown in Figure 18.

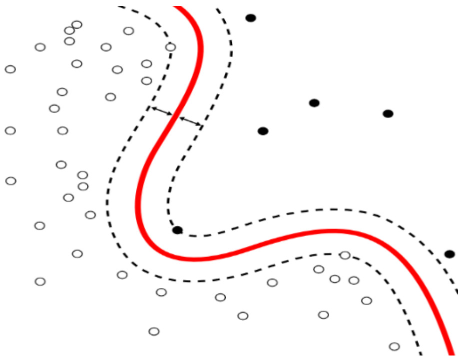


Figure 18. Separation in soft margin [12]

To convert the line to a curve, we define a kernel trick that depicts x on the set of z. According to relations (20):

$$\varphi : x \to z \qquad z = \varphi(x) \tag{20}$$

In other words, relation (21) indicates this:

$$w^T x + b = 0 \to w^T \varphi(x) + b = 0 \to w^T z + b = 0 \tag{21}$$

and in general, the problem has been shown by relations (22), (23) and (24):

$$\min \ \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j y_i y_j \varphi(x_i)^T \varphi(x_j) - \sum_i \alpha_i \tag{22}$$

$$\varphi(x_i)^T \varphi(x_j) = k(x_i, x_j) \tag{23}$$

$$b = \mathrm{mean}(y_i - (\sum_i \alpha_i y_i k(x_i, x))) \tag{24}$$

Kernel trick types are introduced in relations (25), (26) and (27):

1. Polynomials: $k(x_i, x_j) = (1 + x_i{}^T x_j)^P$ (25)

2. Gaussian: $k(x_i, x_j) = \exp(\frac{-1}{2\sigma^2}) \left\| x_i - x_j \right\|^2$ (26)

3. MLP: $k(x_i, x_j) = \tanh(\beta_0 + \beta_1 x_i{}^T x_j)$ (27)

By applying the kernel trick, the soft-margin SVM relationships will change as described in relationships (28), (29), (30) and (31):

Suppose ξ is the error rate:

$$y_i = 1 \Rightarrow w^T x_i + b + \xi \ge 1 \to w^T x_i + b \ge 1 - \xi \tag{28}$$

$$y_i = -1 \Rightarrow w^T x_i + b - \xi \le -1 \to w^T x_i + b \le -1 + \xi \tag{29}$$

$$\Rightarrow y(w^T x + b) \ge 1 - \xi \quad , \quad \xi \ge 0 \tag{30}$$

$$L = \frac{1}{2} w^T w - C \sum_i \xi_i \tag{31}$$

In the above relation, L must be a minimum and C is the penalty coefficient that minimizes or highlights the error's role

$$L = \frac{1}{2} w^T w + C \sum_i \xi_i - \sum_i \alpha_i [y_i (w^T x_i + b) - 1 + \xi] - \sum_i \mu_i \xi_i$$

where, $\frac{1}{2} w^T w$, $C \sum_i \xi_i$ will be minimum and $\sum_i \alpha_i [y_i (w^T x_i + b) - 1 + \xi]$, $\sum_i \mu_i \xi_i$ will be maximum and according to relation (32), (33) and (34):

$$\frac{\partial L}{\partial w} = w - \sum \alpha_i y_i x_i = 0 \Rightarrow w = \sum \alpha_i y_i x_i \tag{32}$$

$$\frac{\partial L}{\partial b} = \sum \alpha_i y_i = 0 \tag{33}$$

$$\frac{\partial L}{\partial \xi} = C - \alpha_i - \mu_i = 0 \tag{34}$$

So, we have the three components with relations (35):

$$C = \alpha_i + \mu_i \qquad \alpha_i \ge 0 \qquad \mu_i \ge 0 \tag{35}$$

and we get the following results that known as Box Constraint.

$$0 \le \alpha_i \le C$$

$$0 \le \mu_i \le C$$

In fact, Box Constraint is the only difference between soft margin and hard margin, and in the hard margin the coefficient C is infinite.

The SVM method is basically a binary classifier, while most of the issues are with multi-class classifiers. In such cases a multi-class problem can be reduced to several binary problems and by comparing pairs of classes and combining their outputs together, a multi-class problem can be solved.

In this article this method is investigated. The SVM method is scheduled with the flowchart shown in Figure 19 and after simulation the CCR is equal to 94.66%.

3. CONCLUSIONS

In this paper, we investigated classification on skin diseases by using different methods such as KNN, Bayesian method, Neural Networks. Due to the not enough amount of data, and also in due to lack of a complete Gaussian distribution on dataset, KNN and Bayesian method do not provide adequate results.

Then Multi-Layer Perceptron, RBF Neural Networks and SVM have been used and the methods provided almost ideal results and thus can be considered as ideal classifiers.

NOMENCLATURES

1. Acronyms

KNN	K-Nearest Neighbor
CCR	Correct Classification Rate
RBF	Radial Basis Function
SVM	Support Vector Machine

2. Symbols / Parameters

P: Data probability
σ: Variance of each data
μ: Average of each data
W: Weight
b : Distance from the origin
L : Saddle equation
α_i : Lagrange coefficient
μ_i : Lagrange coefficient
ξ : Error rate
C : Penalty coefficient

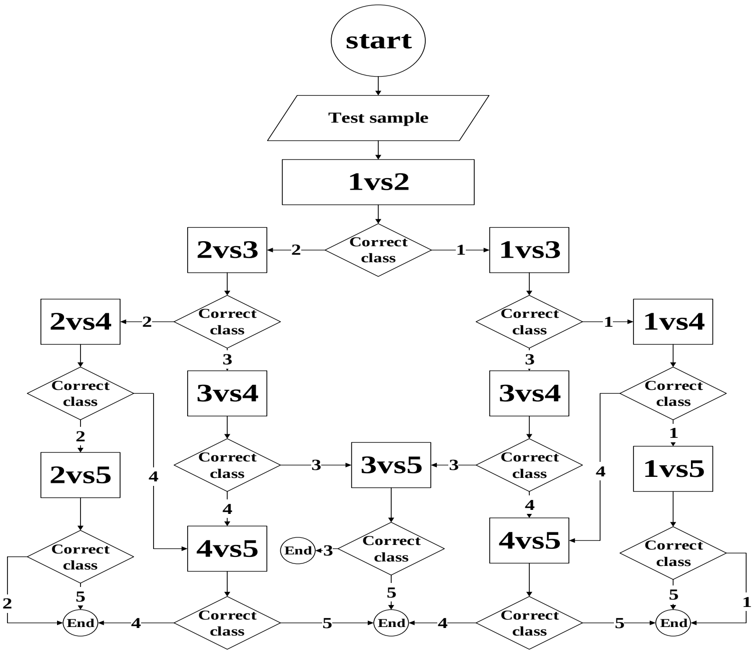


Figure 19. Flowchart for programming the SVM multi-class method

REFERENCES

[1] A.B. Smulevich, A.N. Lvov, O.L. Ivanov, "Self-Inflicted Disorders: Psychopathology of Autoagression in Dermatological Practice", Moscow, Russia, 2012.
[2] J. Todd Williams, "Asheboro Dermatology and Skin Surgery Explains the Acne Treatment", 2016.
[3] H.A. Govenir, N. Ilterand Demiroz, "Learning Differential Diagnosis of Eryhemato-Squamous Diseases using Voting Feature Intervals", Artificial Intelligence in Medicine, Vol. 13, pp. 147-165, 1998.
[4] Y. Lihuna, D. Qi, G. Yanjun, "Study on KNN Text Categorization Algorithm", Micro Computer Information, No. 21, pp. 269-271, 2006.
[5] J. Stallkamp, M. Schlepping, J. Salmen, C. Igel, "The German Traffic Sign Recognition Benchmark: A Multi-Class Classification Competition", International Joint Conference on Neural Networks, pp. 1453-1460, 2011.
[6] Y. Bengio, A. Courville, P. Vincent, "Representation Learning: A Review and New Perspectives", IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1798-1828, 2013.
[7] X. Wang, Y. Wu, Q. Li, T.W. Chan, L. Zhang, S. Wu, "Prediction of the Stress Relaxation Property of Diene Rubber Composites by Artificial Neural Network Approaches", 2015.
[8] J. Park, J. Wsandberg, "Universal Approximation Using Radial Basis Function Network", Neural Comput., Vol. 3, pp. 246-257, 1991.
[9] V.P. Yee, S. Haykin, "Regularized Radial Basis Function Networks and Theory and Applications", John Wiley, 2001.
[10] M. Deshpande, "Using Neural Networks for Regression: Radial Basis Function Networks", 2017.
[11] S. Kampakis, "The Differences between Machine Learning and Statistics", December 5, 2015.
[12] wikipedia.org/wiki/Support-vector_machine.

BIOGRAPHIES

Vahid Yousefi was born in Tabriz, Iran in October 1995. He received the Bachelor degree and graduated in Medical Engineering, Bioelectric Orientation from Tabriz Branch, Islamic Azad University, Tabriz, Iran in 2018. He is currently studying in a Master degree in Medical Engineering, Bioelectric orientation at the same university. His interest in research on classification, data mining, KNN and Bayesian methods, neural network and microcontroller.

Sina Kheiri was born in Tabriz, Iran in September 1989. He received the Bachelor degree and graduated in Medical Engineering, Bioelectric Orientation from Tabriz Branch, Islamic Azad University, Tabriz, Iran in 2013. He is currently studying in a Masters degree in Medical Engineering, Bioelectric Orientation at the same university. His interest in research on classification, data mining, KNN and Bayesian methods, neural network and programming.

Saman Rajebi was born in Tabriz, Iran in 1981. He received his B.Sc. degree in 2003 in Electrical Engineering and M.Sc. and Ph.D. degrees in communication engineering in 2006 and 2017, respectively. He is fulltime teaching staff of Seraj Higher Education Institute, Tabriz, Iran and is member of Executive Committee of ICTPE Conference. His interests are radio frequency electronics, Microwave applications, antenna designing and neural networks.