

Applying Support Vector Machines with Different Kernel to Breast Cancer Diagnosis

Seyyid Ahmed-Medjahed^{1,*}, Fatima Boukhatem²

¹ University of Relizane, Relizane,
Algeria

² University of Djillali Liabes, Sidi Bel Abbes,
Algeria

seyyidahmed.medjahed@univ-relizane.dz, fatima.boukhatem@univ-sba.dz

Abstract. The detection of breast cancer poses a significant challenge in the field of medicine. It represents the second type of the largest cases of cancer deaths in women. Several techniques have been found to solve the problem or make a better diagnosis. Recently, Support Vector Machine based systems are the most common and are considered a better diagnostic assistant in cancer detection research. The quality of the results generated depends on the choice of some parameters such as the kernel function and the model parameters. In this paper, we analyze and evaluate the performance of several kernel functions in the SVM algorithm. Experiments are conducted with different training-test phases generated by the holdout method and we used the WBCD (Wisconsin Breast Cancer Database) to analyze the results. The results are evaluated by using the following performances measures: classification accuracy rate, sensitivity, specificity, positive and negative predictive values. To validate the results obtained by these different kernel functions, we use different values for the kernel functions parameters and SVM model parameters and we record the optimal parameters values. Finally, we show that the Cauchy kernel and the Rational Quadratic kernel are identical and converge to the same value.

Keywords. Support vector machine, kernel function, breast cancer, diagnosis, classification, sequential minimal optimization.

1 Introduction

Breast cancer originates from the inner lining of milk ducts or the lobules responsible for supplying

milk to the ducts. It manifests as a tumor within the breast, which can either be benign (non-cancerous) or malignant (cancerous).

Malignant tumors grow and develop into cancer. Breast cancer is a leading cause of mortality worldwide. Every year the breast cancer is detected in 1,3 million women on the world, the mortality rate increases rapidly with more than 1,6 million cases in 2010 which corresponds to 425000 deaths [6]. The correct diagnosis of breast cancer is very important to help the doctors.

Detecting cancerous cells at an early stage, before they spread, can significantly increase the survival rate for patients by over 97%¹ Nevertheless, classifier systems are widely used to solve the problem of cancer classification and to help the experts to make a good diagnosis. The major advantages of classifier systems are the minimization of possible errors that might be made and the ability to give a detailed examination.

The Support Vector Machine (SVM) is recognized as one of the most widely used classifier systems which has become quite an active research field in machine learning. Recently, Support Vector Machines have been shown to give good results and a good generalization performance in the medical diagnosis field precisely in cancer classification [19, 8, 11, 3, 17].

¹American Cancer Society Homepage. (2008). Citing Internet sources Available from: www.cancer.org.

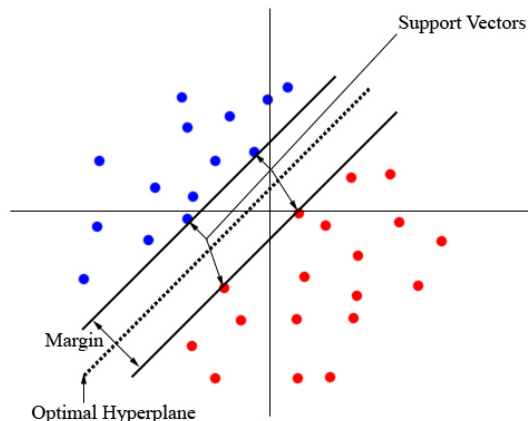


Fig. 1. The figure shows the optimal hyperplane and the support vectors

Table 1. Some kernel functions defined in the literature (M. is multiquadric and q. is quadratic)

Kernel name	Formulation
Linear	$k(x, y) = x^t \cdot y$
Polynomial	$k(x, y) = (x^t \cdot y)^d$
Gaussian	$k(x, y) = \exp(-\frac{\ x-y\ ^2}{2\sigma^2})$
Sigmoid	$k(x, y) = \tan(P1 \cdot x^t \cdot y + P2)$
Cauchy	$k(x, y) = \frac{1}{1 + \frac{\ x-y\ ^2}{\sigma^2}}$
Inverse M.	$k(x, y) = \frac{1}{\sqrt{\ x-y\ ^2 + c^2}}$
Quadratic	$k(x, y) = (x^t \cdot y + 1)^2$
Multiquadric	$k(x, y) = \sqrt{\ x-y\ ^2 + c^2}$
Power	$k(x, y) = -\ x-y\ ^d$
Rational Q.	$k(x, y) = 1 - \frac{\ x-y\ ^2}{\ x-y\ ^2 + c}$
Wave	$k(x, y) = \frac{\theta}{\ x-y\ } \sin \frac{\ x-y\ }{\theta}$
Spherical	$k(x, y) = 1 - \frac{3}{2} \frac{\ x-y\ }{\sigma} + \frac{1}{2} (\frac{\ x-y\ }{\sigma})^3$

Achieving strong performance with the SVM method heavily relies on the selection of appropriate kernel functions. These functions enable the algorithm to identify the maximum-margin hyperplane within a transformed feature space.

Additionally, the effectiveness of the SVM hinges on the careful tuning of kernel function

parameters and the cost parameter C in the SVM model has a very important role to get a good classification accuracy rate [18]. Numerous studies in the literature have explored medical diagnosis of breast cancer using the Wisconsin Breast Cancer Database (WBCD).

For instance, Quinlan achieved a classification accuracy rate of 94,74% through 10-fold cross-validation using the C4.5 decision tree method [16]. Similarly, Hamilton et al. achieved a classification accuracy rate of 95,00% using the RIAC method [9].

Nauck and Kruse obtained 95,06% classification accuracy rate by using the neuron-fuzzy techniques [13]. Albrecht et al. reached 98,80% classification accuracy rate with logarithmic simulated annealing with the perceptron algorithm [2].

Übeyli, by using SVM reached 99,54% accuracy [21]. Polat and Günes used the LS-SVM (Lest Square SVM) and 98,53% was obtained [15]. Guijarro-Berdias et al. achieved 96,00% classification accuracy rate by applying linear-least squares [7].

Akay, by using SVM with feature selection, reached 99,51% classification accuracy rate [1]. Marcano-Cedeño et al., by applying artificial metaplasticity neural network reached 99.26% classification accuracy rate [12]. In this paper, we compare and analyze several kernel functions proposed in the literature by using different values of kernel functions parameters and cost parameters.

This study has been applied to the Wisconsin Breast Cancer Dataset (WBCD) which is a widely studied data set from the field of breast cancer diagnosis. We evaluate the results by calculating the performance measures: classification accuracy rate, sensitivity, specificity, positive and negative predictive values.

In this work we have done two studies: the first one is the realization of a comparison protocol between the different kernels and the second is to reach a high classification accuracy rate in the context of breast cancer diagnosis.

Also, we show that the Cauchy kernel and the Rational Quadratic kernel are identical and give the same results.

Table 2. Kernel functions with parameters

Kernel name	Kernel Parameter
Linear	/
Polynomial	d
Gaussian	σ
Sigmoid	P_1, P_2
Cauchy	σ
Inverse Multiquadric	c
Quadratic	/
Multiquadric	c
Power	d
Rational Quadratic	c
Wave	θ
Spherical	σ

Table 3. WBCD description of attributes

Attribute numbers	Attribute description	Values of attribute
1	Clump thickness	1-10
2	Uniformity of cell size	1-10
3	Uniformity of cell shape	1-10
4	Marginal adhesion	1-10
5	Single epithelial cell size	1-10
6	Bare nuclei	1-10
7	Bland chromatin	1-10
8	Normal nucleoli	1-10
9	Mitoses	1-10

The rest of the paper is organized as follows: First in Section 2, we give an overview of SVM. In Section 3, we recall some kernel functions defined in the literature. In Section 4, we analyze the results of the different kernel function. Finally, we conclude with some perspectives.

2 Overview of Support Vector Machine

Vapnik [4] introduced the Support Vector Machine (SVM) as a learning algorithm aimed at minimizing structural risk. SVM is a method used for data

analysis and is employed in both classification and regression tasks.

Given a set of input data, SVM predicts which of two possible classes each input belongs to. To achieve classification, SVM constructs a hyperplane in a high-dimensional space to effectively separate the data into classes.

This hyperplane is positioned to maximize the distance to the nearest training point of any class, ensuring optimal separation. While multiple valid hyperplanes exist, SVM uniquely identifies the optimal hyperplane. The data points that are closest to this maximum margin hyperplane are referred to as Support Vectors.

Identifying this hyperplane involves reformulating the classification problem into a quadratic optimization task, which can be resolved using various algorithms such as Sequential Minimal Optimization, Trust Region, Interior Point, Active-Set, and others. One key benefit of SVM is its effectiveness in high-dimensional spaces, even when the number of dimensions exceeds the number of samples.

2.1 Mathematical Formulation

Given a training data set of N points (x_i, y_i) with input data $x_i \in \mathbb{R}^d$, $i = 1, \dots, N$ and output data $y_i \in \{-1, 1\}$ given by an expert. The margin is the distance of closest examples from the line decision (hyperplane). The equation of hyperplane can be written as the set of points X satisfy:

$$\langle w, x_i \rangle + b = 0. \quad (1)$$

The hyperplane that optimally separates the data is the one that minimizes $:\frac{1}{2}w^T w$.

This gives the final standard formulation of an SVM as a minimization problem:

$$\begin{cases} \min \frac{1}{2}w^T w, \\ y_i(\langle w_i, x_i \rangle + b) \geq 1, i = 1, \dots, N. \end{cases} \quad (2)$$

This represents a quadratic programming optimization challenge. Quadratic optimization problems are a widely recognized category of mathematical optimization problems, with numerous algorithms available for their resolution.

Table 4. The results obtained by different kernel functions with the bestvalue of kernel parameter

Kernel name	Results	Kernel Parameter
Linear	97.40	l
Polynomial	96.54	$d = 3$
Gaussian	99.13	$\sigma = 3$
Sigmoid	96.53	$P_1 = 1, P_2 = -1$
Cauchy	98.27	$\sigma = 10$
Inverse Multiquadric	98.27	$c = 1$
Quadratic	93.59	$c = 5$
Multiquadric	64.93	$c = 9$
Power	34.63	$d = 2$
Rational Quadratic	98.27	$\sigma = 10$
Wave	97.83	$\theta = 2$
Spherical	64.93	$\sigma = 0.1$

The dual problem is obtained by introducing Lagrange multipliers:

$$\begin{cases} \max \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \langle x_i, x_j \rangle, \\ \alpha_i \geq 0, \\ \sum_{i=1}^N \alpha_i y_i = 0. \end{cases} \quad (3)$$

Solving equation (3) with constraints equation determines the lagrange multipliers, and the optimal separating hyperplane is given by:

$$w^* = \sum_{i=1}^N \alpha_i y_i x_i, \quad (4)$$

$$b^* = -\frac{1}{2} \langle w^*, x_r + x_s \rangle, \quad (5)$$

where x_r and x_s are any support vector from each class satisfying:

$$\alpha_r, \alpha_s > 0, y_r = -1, y_s = 1. \quad (6)$$

The hard classifier is then:

$$f(x) = \text{sign}(\langle w^*, x \rangle + b^*). \quad (7)$$

In this study, we choose the Sequential Minimal Optimization algorithm to solve the quadratic problem.

2.2 Sequential Minimal Optimization

The Sequential Minimal Optimization (SMO), introduced by [10, 14] is another widely used algorithm for training Support Vector Machines (SVMs). The basic idea of smo is to decompose the initial problem into sub problems (reducing the working sets to two points).

the optimal solution can be computed analytically for this two points in the working set [5]. Given the current solution $(\alpha_i^{\text{old}}, \alpha_j^{\text{old}})$, the optimal update is computed to obtain the new solution $(\alpha_i^{\text{new}}, \alpha_j^{\text{new}})$ by using the following update rule:

$$\alpha_j^{\text{new}} = \alpha_j^{\text{old}} - \frac{y_j(E_i - E_j)}{\eta}, \quad (8)$$

where

$$E_k = f(x_k) - y_k, \quad (9)$$

$$\eta = 2\langle x_i, x_j \rangle - \langle x_i, x_i \rangle - \langle x_j, x_j \rangle, \quad (10)$$

where E_k is the error between the SVM output on the k^{th} example and the true label y_k .

Next we clip α_j^{new} to lie within the range $[L, H]$ i.e $L \leq \alpha_j^{\text{new}} \leq H$, to satisfy the constraint that $0 \leq \alpha_j \leq C$:

$$\alpha_j^{\text{new}} = \begin{cases} H & \text{si } \alpha_j^{\text{new}} \geq H, \\ L & \text{si } \alpha_j^{\text{new}} \leq L, \\ \alpha_j^{\text{new}} & \text{si } L < \alpha_j^{\text{new}} < H. \end{cases} \quad (11)$$

The bounds L and H are given by the following:

– If $y_i \neq y_j$, $L = \max(0, \alpha_j^{\text{old}} - \alpha_i^{\text{old}})$, $H = \min(C, C + \alpha_j^{\text{old}} - \alpha_i^{\text{old}})$

– If $y_i = y_j$, $L = \max(0, \alpha_i^{\text{old}} + \alpha_j^{\text{old}} - C)$, $H = \min(C, \alpha_i^{\text{old}} + \alpha_j^{\text{old}})$.

Finally, having solved for α_j^{new} , the value of α_i^{new} is given by:

$$\alpha_i^{\text{new}} = \alpha_i^{\text{old}} + y_i y_j (\alpha_i^{\text{old}} - \alpha_j^{\text{old}}). \quad (12)$$

The algorithm proceeds as follows:

1. Find a Lagrange multiplier α_1 that violates the Karush – Kuhn – Tucker (KKT) conditions for the optimization problem.

Table 5. The results obtained by different kernel functions with the bestvalue of SVM model parameter

Kernel name	cost parameter C
Linear	1
Polynomial	1
Gaussian	13
Sigmoid	1
Cauchy	6
Inverse Multiquadric	2
Quadratic	5
Multiquadric	1
Power	2
Rational Quadratic	6
Wave	1
Spherical	1

Table 6. Sensitivity and specificity calculated for the four best kernel functions

Kernel name	Sensitivity	Specificity
Gaussian	98.67	100
Inverse M.	98.67	97.53
Cauchy	98.67	97.53
Wave	98.67	96.30

- Pick a second multiplier α_2 and optimize the pair (α_1, α_2) .
- Repeat steps 1 and 2 until convergence.

Upon satisfaction of the Karush-Kuhn-Tucker (KKT) conditions by all Lagrange multipliers within a specified user-defined tolerance, the problem is considered solved.

While this algorithm ensures convergence, heuristics are employed to select the pair of multipliers to expedite convergence. To achieve optimal performance, certain parameters in SVM must be meticulously chosen. These parameters include:

- The regularization parameter C , which controls the trade-off between errors of the SVM on training data and margin maximization [20].

- The parameters of the kernel functions.
- The choice of the kernel affect the performance.

3 Kernel Functions

However, in 1992, V. Vapnik et al. proposed a method to generate nonlinear classifiers by employing the kernel trick with maximum-margin hyperplanes.

Kernel functions enable a nonlinear transformation of data into a linear separation of examples in a new space known as the "feature space," which is high-dimensional.

This characteristic enhances the likelihood of discovering a separating hyperplane. Nevertheless, in this new space, the goal is to find the following hyperplane:

$$h(x) = \langle w, \Phi(x) \rangle + b. \quad (13)$$

We arrive at the following optimization problem:

$$\begin{cases} \max \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \langle \Phi(x_i), \Phi(x_j) \rangle, \\ \alpha_i \geq 0, \\ \sum_{i=1}^N \alpha_i y_i = 0. \end{cases} \quad (14)$$

By introducing the notion of kernel function we have:

$$k(x_i, x_j) = \Phi(x_i) \Phi(x_j). \quad (15)$$

The expression of the hyperplane will be defined as follows:

$$f(x) = \sum_i \alpha_i y_i \Phi(x_i) \Phi(x_j) + b. \quad (16)$$

Under these conditions we didn't need to know the transformation Φ and the calculation will be much less expensive.

We can directly construct a kernel function by respecting certain condition defined by the theorem of Mercer which states that: a kernel $k(x_i, x_j)$ is a symmetric continuous function that maps two variables to a real value and k must be positive semi definite.

Unfortunately, this theoretical condition is difficult to verify, it does not provide guidance for the construction of kernels or on the transformation

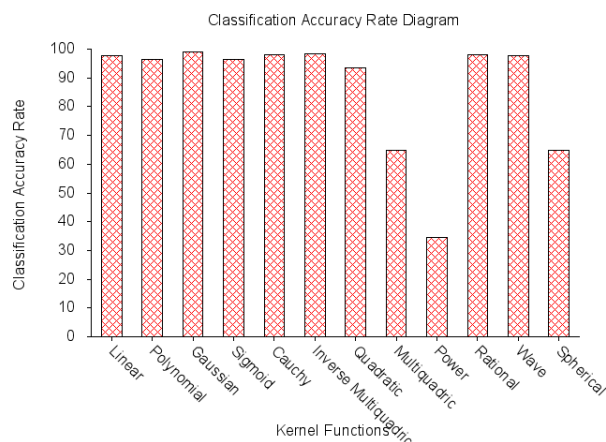


Fig. 2. The figure shows the result obtained by the different kernel

Φ . Several researches has been devoted to constructing a kernel more exotic and adapted to a special problem. In table 2, we show the several kernel functions proposed in the literature:

4 Experimentation

In this research, we analyze and evaluate the performance of different kernel functions in the SVM algorithm by using different values of kernel functions parameters and SVM model parameters. We use the SMO algorithm to solve the quadratic problem of maximizing the margin.

This analysis is conducted using the publicly accessible breast cancer database known as WBCD (Wisconsin Breast Cancer Database), which originates from the work undertaken at the University of Wisconsin Hospital.

This set of data was taken from Fine Needle Aspirates (FNA) of human breast tissue classified as benign and malignant. The WBCD database contains 699 clinical cases, there is 458 (65,50%) benign cases and 241 (34,50%) malignant cases.

The data of WBCD contains 16 instances with missing attribute values which led us to limit our experimentation to 683 clinical cases. Nevertheless, the class has a distribution of 444 (65%) benign cases and 239 (35%) malignant cases. Each instance in the database has a nine

Table 7. Positive predictive value and negative predictive value calculated for the four best kernel functions

Kernel name	Pos. Pre. Val.	Neg. Pre. Val.
Gaussian	100	97.59
Inverse M.	98.67	97.53
Cauchy	98.67	97.53
Wave	98.01	97.50

attributes; each attributes has an integer value between 1 and 10.

In the following table, we detailed the attributes of WBCD: The SVM method consists of two phases: training and testing. To randomly divide the database into two parts, we employ the holdout method, a form of cross-validation.

This method randomly partitions the initial data into two sets: the training set and the testing set. Less than one-third of the initial data is allocated for testing purposes. With the holdout method, we obtained 455 samples (65,10%) for the training phase and 244 samples (34,90%) for the testing phase.

In the first step, we analyze and compare the results obtained by the different kernel functions in terms of accuracy classification rate. The values of kernel functions parameters were chosen by experimentation.

For the cost parameter C (regularization parameters that control the flexibility), we have varied its value between 0.01 and 1000 then we record the values which give a good accuracy classification rate.

In the second step, we take the four kernel functions which have given a good results and we analyze these kernels in the function of: sensitivity, specificity, positive and negative predictive values. The parameters measure are calculated using the following equations:

where:

NTP : Number of True Positives

NTN : Number of True Negatives

NFP : Number of False Positives

NFN : Number of False Negatives

Table 8. Classification accuracy rate (CAR) obtained with SVM by using the Gaussian kernel and other classifiers

Author and years	CAR
Quinlan (1996)	94.74
Hamiton et al. (1996)	95.00
Nauck and Kruse (1999)	95.06
Albrecht et al. (2002)	98.80
Polat and Günes (2007)	98.53
Guijarro-Berdias et al. (2007)	96.00
Akay (2009)	99.51
A. Marcano-Cedeño (2011)	99.26
<i>This Study</i>	99.13

Classification accuracy:	$\frac{NTP + NTN}{NTP + NTN + NFP + NFN}$
Sensitivity:	$\frac{NTP}{NTP + NFN}$
Specificity:	$\frac{NTN}{NFP + NTN}$
Positive Predictive Value:	$\frac{NTP}{NTP + NFP}$
Negative Predictive Value:	$\frac{NTN}{NTN + NFN}$

In table 5, we present the good results (the high value of: classification accuracy rate and its kernel parameters) obtained by each kernel function after several trials and evaluation of different values of the kernel functions parameters with SVM model parameters. We clearly observe, that the Gaussian, Cauchy, inverse multiquadratic, rational quadratic, linear and polynomial kernels have given a good results with an advantage for Gaussian kernel function (99,13%).

We show also, that the Cauchy, inverse multiquadratic and rational quadratic kernel have nearly the same classification accuracy rate. The low classification accuracy rate is registered for the multiquadratic and spherical kernel function with 64,93%.

During the assessments, we show that the results obtained by the rational quadratic kernel and the Cauchy kernel are identical despite the different tests with different training and testing set.

Therefore, we will show that the rational quadratic kernel and the Cauchy kernel are identical and converge to the same value:

$$\begin{aligned}
 k(x, y) &= \frac{1}{1 + \frac{\|x - y\|^2}{c}}, \\
 &= \frac{1}{\frac{c + \|x - y\|^2}{c}}, \\
 &= \frac{c}{c + \|x - y\|^2}.
 \end{aligned} \tag{17}$$

The Rational Quadratic kernel is defined as:

$$\begin{aligned}
 k(x, y) &= 1 - \frac{\|x - y\|^2}{\|x - y\|^2 + c}, \\
 &= \frac{\|x - y\|^2 + c}{\|x - y\|^2 + c} - \frac{\|x - y\|^2}{\|x - y\|^2 + c}, \\
 \text{with } 1 &= \frac{\|x - y\|^2 + c}{\|x - y\|^2 + c}, \\
 &= \frac{c}{\|x - y\|^2 + c}.
 \end{aligned} \tag{18}$$

Finally, we can say that using the rational quadratic kernel function gives the same results as the Cauchy kernel function. So, the transformation of data by these kernel are the same and give the identical new space. In table 7, we show: sensitivity, specificity, positive and negative predictive values obtained by the four kernels which have given a high diagnostic accuracy.

Table 8, gives the classification accuracies of SVM with Gaussian kernel and previous methods applied to the same database.

5 Conclusion

In this study, we have analyze and compared the performance of several kernel on the support vector machine in the context of breast cancer diagnosis. We conducted our experimentation on the WBCD database. We have shown that the Gaussian kernel has given a good results in term of classification accuracy rate (99,13%). The Gaussian kernel have a great performance with: 98,67% in sensitivity and 100% in specificity. Also, we have shown that the both Cauchy kernel and the Rational Quadratic kernel are the same and give the same results.

References

1. **Akay, M. F. (2009).** Support vector machines combined with feature selection for breast cancer diagnosis. *Expert Systems with Applications*, Vol. 36, No. 2, pp. 3240–3247. DOI: 10.1016/j.eswa.2008.01.009.
2. **Albrecht, A. A., Lappas, G., Vinterbo, S. A., Wong, C. K., Ohno-Machado, L. (2002).** Two applications of the Isa machine. *Proceedings of the 9th international conference on neural information processing*, pp. 184–189. DOI: 10.1109/ICONIP.2002.1202156.
3. **Brook, A., El-Yaniv, R., Isler, E., Kimmel, R., Meir, R., Peleg, D. (2006).** Breast cancer diagnosis from biopsy images using generic features and SVMs. *IEEE Transactions on Information Technology in Biomedicine*.
4. **Cortes, C., Vapnik, V. (1995).** Support-vector networks. *Machine Learning*, Vol. 20, pp. 273–297. DOI: 10.1007/Bf00994018.
5. **Flake, G. W., Lawrence, S. (2002).** Efficient SVM regression training with SMO. *Machine Learning*, Vol. 46, pp. 271–290. DOI: 10.1023/A:1012474916001.
6. **Forouzanfar, M. H., Foreman, K. J., Delossantos, A. M., Lozano, R., Lopez, A. D., Murray, C. J., Naghavi, M. (2011).** Breast and cervical cancer in 187 countries between 1980 and 2010: A systematic analysis. *The Lancet*, Vol. 378, No. 9801.
7. **Guijarro-Berdias, B., Fontenla-Romero, O., Perez-Sanchez, B., Fraguera, P. (2007).** A linear learning method for multilayer perceptrons using leastsquares. *Lecture Notes in Computer Science*, pp. 365–374. DOI: 10.1007/978-3-540-77226-2_38.
8. **Guyon, I., Weston, J., Barnhill, S., Vapnik, V. (2002).** Gene selection for cancer classification using support vector machines. *Machine Learning*, Vol. 46, pp. 389–422. DOI: 10.1023/A:1012487302797.
9. **Hamilton, H. J., Shan, N., Cercone, N. (1996).** A rule induction algorithm based on approximate classification. *Journal of Computer and System Science*, Vol. 20, No. 11, pp. 34–50.
10. **Keerthi, S. S., Shevade, S. K., Bhattacharyya, C., Murthy, K. R. K. (2001).** Improvements to platt's SMO algorithm for SVM classifier design. *Neural Computation*, Vol. 13, No. 3. DOI: 10.1162/089976601300014493.
11. **Mallika, R., Saravanan, V. (2010).** An SVM based classification method for cancer data using minimum microarray gene expressions. *International Journal of Computer and Information Engineering*, Vol. 4, No. 2, pp. 266–270.
12. **Marcano-Cedeno, A., Quintanilla-Domínguez, J., Andina, D. (2011).** WBCD breast cancer database classification applying artificial metaplasticity neural network. *Expert Systems with Applications*, Vol. 38, No. 8, pp. 9573–9579. DOI: 10.1016/j.eswa.2011.01.167.
13. **Nauck, D., Kruse, R. (1999).** Obtaining interpretable fuzzy classification rules from medical data. *Artificial Intelligence in Medicine*, Vol. 16, No. 2. DOI: 10.1016/S0933-3657(98)00070-0.
14. **Platt, J. C. (1999).** Fast training of support vector machines using sequential minimal optimization. *Advances in Kernel Methods - Support Vector Learning*. DOI: 10.7551/mitpress/1130.003.0016.
15. **Polat, K., Günes, S. (2007).** Breast cancer diagnosis using least square support vector machine. *Digital Signal Processing*, Vol. 17, No. 4, pp. 694–701. DOI: 10.1016/j.dsp.2006.10.008.
16. **Quinlan, J. R. (1996).** Improved use of continuous attributes in C4.5. *Journal of Artificial Intelligence Research*, Vol. 4, pp. 77–90. DOI: 10.1613/jair.279.
17. **Rejani, Y. I. A., Selvi, S. T. (2009).** Early detection of breast cancer using SVM classifier technique. *International Journal on Computer*

Science and Engineering, Vol. 1, No. 3.
DOI: 10.48550/arXiv.0912.2314.

18. **Rychetsky, M. (2001).** Algorithms and architectures for machine learning based on regularized neural networks and support vector approaches. Shaker, Germany.
19. **Shah, S., Kusiak, A. (2002).** Cancer gene search with data-mining and genetic algorithms. *Computers in Biology and Medicine*, Vol. 37, No. 2, pp. 251–261. DOI: 10.1016/j.compbiomed.2006.01.007.

20. **Shawe-Taylor, J., Cristianini, N. (2004).** *Kernel Methods for Pattern Analysis*. Cambridge University Press.

21. **Übeyli, E. D. (2007).** Implementing automated diagnostic systems for breast cancer detection. *Expert Systems with Applications*, Vol. 33, No. 4, pp. 1054–1062. DOI: 10.1016/j.eswa.2006.08.005.

Article received on 19/12/2023; accepted on 18/04/2024.
** Corresponding author is Seyyid Ahmed-Medjahed.*