A MACHINE LEARNING MODEL FOR BREAST CANCER DETECTIONUSING SUPPORT VECTOR MACHINE

BURHANUDDIN QURESHI[1] | HASSAN Md [2] | MUHAMMED HABEEB[3] | S.G.M. SHAADAB[4]

[1,2,3,4]Lecturer, Department of Information Technology, University of Technology and Applied Science AlMusanna, Sultanate of Oman

Abstract

It investigates the reduction of features in a breast cancer decision assistance system. The WDBC dataset is converted to a one-dimensional feature vector (IC). The original data with 30 features and one reduced feature (IC) is used to evaluate diagnostic accuracy of classifier using support vector machine (SVM). The suggested classification using the IC is compared to the original feature set usingdifferent validation (5/10-fold cross-validations) and partitioning (20%–40%) methods. Its performance is measured in terms of specificity, sensitivity, accuracy, F-score, You den's index, discriminant power,and the receiver operating characteristic (ROC) curve with its criterion values of area under curve (AUC) and 95% confidence interval (CI) (CI). This reduces computing complexity while improving diagnostic decision assistance.

Keywords: Support vector machine, machine learning, classification, WDBC datasets.

1. Introduction

In women, cancers of the breast are among the most common death causes from all types of cancer[1]. The best way to treat cancer is to find it early and get the right diagnosis. However, traditional cancer diagnosis relies on doctors' experience and their ability to look at the body.

It is normal for humans to make mistakes because they have so many things they can't do. Humans are good at noticing patterns, but they don't know how likely they are to happen [2]. Even with a lot of tests, it can be hard to get a precise diagnosis even for experts That's why a lot of researchers are looking towards automaticdetection of breast cancer. Doctors can use computer-aided diagnostic technologies to increase their accuracy in making diagnosis. [3-5].

It turns out that the most experienced doctor can make 79.97% of correct diagnoses, but with the help of machine learning it is 91.1 percent correct[6].

It is dangerous or cancerous to have a malignant tumor. Benign tumors, on the other hand, can make it more likely that you'll get breast cancer. People find malignant tumors more alarming than benign tumors, because they are more likely to becancerous than benign tumors, even though a lot of research has been done to find cancerous tumors early, about 20% of all women who have them die from them. Backpropagation artificial neural networks (ANN) were used to see how well they couldclassify breast masses as either benign or malignant [8]. Also, radial basis function neural networks (RBFNN) have been shown to be very accurate at detecting microcalcification [9, 10].

Simple construction, strong performance with nonlinear functions and fast convergence time are some of the advantages of RBFNNs. In this way, it has been used a lot for things like recognizing designs and modeling the system[11,12].

This is not true when it comes to RBFNN. The structure of this network grows as the net's input dimension grows. Another thing that will hurt RBFNN's ability to generalizeis if it has input components that aren't important.

SVM is a good statistical method for classifying things [14]. It works by finding the best hyperplane to move input data into a more detailed feature space. Even if there is a lot of data to work with, SVM can be trained quickly [15, 16]. A wide range of recognition issues, including as object and face detection, have been addressed usingit. [17–19].

Using second-order statistical data, principal component analysis (PCA) can lower thenumber of features in a dataset [20]. Independent component analysis (ICA) is a newtechnique in the domains of pattern recognition and signal processing.

Statistics of high order are used to identify elements with more facts than can be obtained via principal component analysis. It is possible to lower the objects in numbers to learn by employing ICA prior to training SVM. There are ways to cut downon the complexity of classifiers, while speeding up their convergence and performance[13, 23].

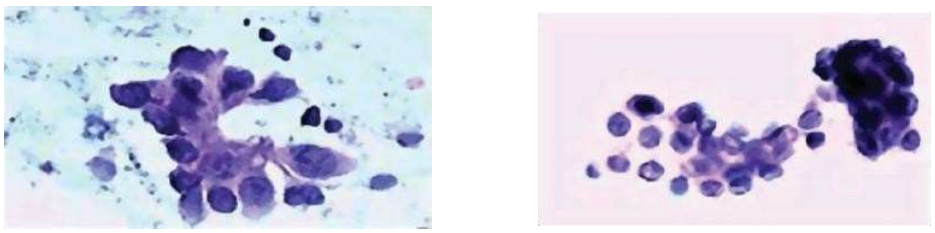
By reducing the number of features, SVM can help us decide if a tumor is benign or malignant, which is the focus of the proposed research. With the use of ICA, the dimensions of the WDBC dataset are decreased to just one. The data that is reducedor lowered is split into test and training data respectively using 10-fold cross-validationand 20 percentage of partitioning in order to assess the accuracy of SVM. The ROC curve Indicators of performance, such as precision, specificity, and sensitivity are all included in this graph, as are F-score, Youden's index, and discriminant power. [42].

ICA, SVM, and performance measurements are summarized in Section 2 of this research. The study's methodology is described in detail in Section 3. Results are discussed in sections 4 and 5 of this chapter. Section 6 concludes with a conclusion.

2. Data sets and SVMs.

Specifications of datasets: 2.1. The WBDC dataset has 569 occurrences, 357 of whichare good and 212 of which are bad. Each sample has a unique identification number,a diagnosis either benign or malignant, and thirty different characteristics. This is howit works: There was a FNA done on this breast lump, and this digital picture was usedto figure out its features.

Each cell nucleus had 30 characteristics. These three metrics were calculated for eachtrait: average, standard error, and "worst" or "worst" (mean of the three greatest numbers).



(a) (b)

Figure 1: Fine needle analysis tissue of breast. Malignant (a) and benign (b)breast tumors [24].

The 10 features worth mentioning are radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry and fractal dimension

It is referred to as the Support Vector Machine (SVM). SVM is a sort of supervised learning that is used in this case. It has been used to categorize and predict data. Theconcept was originally proposed by Boser et al. [30] and Vapnik [31]. A hyperplane that separates the classes is discovered using the SVM approach, which keeps the training error and margin low while maintaining a large margin of safety.

In order to classify datasets that can be separated in a linear fashion, a linear SVM algorithm can be utilized. It is the algorithm's goal to increase the margin as much as is reasonably achievable. Figure 2 depicts support vectors, which are points on the edges that demonstrate how they assist the body in maintaining its strength.

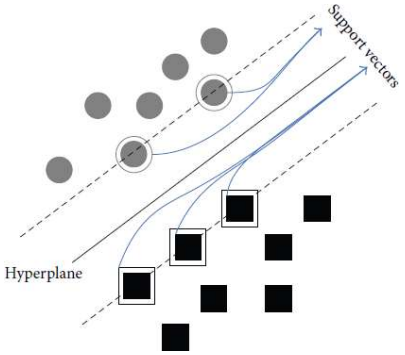


Figure 2: The separating hyperplane with support vectors.

When data points are represented by x, a coefficient vector is represented by w, and an offset from the origin is represented by b, the discriminant function of the hyperplane is given by the equation g (x) = wTx + b (1) In linear SVM, there are two cases,

Where g(x) is greater than equal to 0 for the closest point on one of the classes, andg(x) less than 0 for the closest point on another class.

While maintaining the cost function as low as possible, the margin (2/||w||2) should be maximized to increase generalization ability while keeping the cost function as low aspossible:

$J(w)=1\div2|\|w\|^2|$

(2)

$y_i(W^T x_i + b) \geq 1 \; i = 1, 2, \ldots, n$ and $y_i = \{+1, -1\}$ denotes class labels.

With respect to a set of linear inequality constraints, this is a quadratic optimization problem. The Lagrange function is discovered using Karush-Kuhn-Tucker (KKT)conditions.

$$L_p(w, b, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n}\alpha_i \left\{y_i\left(W^T x_i + b\right) - 1\right\}$$

(3)

L_p must be minimized to get the optimal w and b, where α_i are Lagrange multipliers. The optimization equation is as follows:

$$\text{Maximize}\left[\sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j x_i^T x_j\right] \quad (4)$$

SVM can also be used to address nonlinear classification issues using the kernel function approach, which is another use of the technology. This function transforms data points into a higher-dimensional space in order to generate a hyperplane separating the classes. It has been revealed that a new discriminant function, as shown below, can be used to identify

(5)

$$g(x) = W^T \Phi(X) + b,$$

where (X) denotes the mapping of input vectors to the output vectors. As a result, thefollowing is an example of how to write the optimization equation:

$$\text{Maximize}\left[\sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j K(x_i, x_j)\right] \quad (6)$$

Table 1: Confusion Matrix.

Actual	Observed value	
	Positive values	Negative values
Positive(+)	Tpositive	Fpositive
Negative(-)	Fpositive	Tnegative

2.3 *Performance Measures.*

It is possible to evaluate the performance of classifiers in a variety of methods. To assess the classifier's quality, a confusion matrix is used to keep track of correct and wrong classification results. True positive, true negative, false positive, and false

negative numbers are represented in the binary classification confusion matrix shownin Table 2, with the letters Tpositive, Tnegative, and Fpositive denoting true positives,true negatives, and false positives denoting false negatives, respectively.

The accuracy for classifier is the most frequent empirical measure for assessing efficacy, and it is calculated by

$$\text{Accuracy} = \frac{\text{Tpositive} + \text{Tnegative}}{\text{Tpositive+Tnegative+Fpositive+Fnegative}} \quad (7)$$

The percentage of true positives that are correctly identified is referred to as sensitivity,whereas the percentage of true negatives that are correctly detected is referred to as specificity. This evaluated using below formula

$$\text{Sensitivity} = \frac{\text{Tpositive}}{\text{Tpositive} + \text{Fnegative}}$$

$$\text{Specificity} = \frac{\text{Tpositive}}{\text{Tpositive} + \text{Fnegative}} \quad (8)$$

The F-score is a metric for measuring test accuracy. Both precision and recall are taken into consideration when computing.

$$\text{Precision} = \frac{\text{Tpositive}}{\text{Tpositive} + \text{Fpositive}}$$

$$\text{Recall} = \frac{\text{Tpositive}}{\text{Tpositive} + \text{Fnegative}} \quad (9)$$

$$\text{F-score} = \frac{(\beta^2 + 1) \times \text{precision} \times \text{recall}}{\beta^2 \times \text{precision} + \text{recall}}$$

When = 1, F-Score is balanced, and is the bias. When 1 is present, it encourages recall; otherwise, it prefers precision.

A classifier's discriminant power (DP) and Youden's index are two additional methodsfor evaluating the efficacy of a classifier in medical diagnosis. The capacity of a classifier to discriminate between positive and negative samples is evaluated using the DP method:

$$DP = \frac{\sqrt{3}}{\pi}(\log X + \log Y) \quad (10)$$

$$X = \frac{\text{sensitivity}}{1 - \text{sensitivity}}$$

$$Y = \frac{\text{specificity}}{1 - \text{specificity}} \quad (11)$$

The following is a synopsis of the research findings: In the first case, "poor discriminant," "limited discriminant," "fair discriminant," and "other cases" are all followed by "good discriminant." When it comes to classifiers, Youden's index measures their capacity to prevent failure.[33]

$$\gamma = \text{sensitivity} - (1 - \text{specificity}) \quad (12)$$

The receiver operating characteristic (ROC) curve is summarized using Youden's index. The ROC curve analysis is used to look at how well a test or a classifier can tell the difference between people who are sick and people who are healthy.
Using 5/10-fold cross-validation (CV) and 20 percent data partitioning, the researchers hope to determine how well classifiers that apply the criteria above perform. If you utilize 5-CV or 10-CV, the data is divided into groups of 5 or 10 individuals, and each group is used as test data, with the remaining groups serving as training data. Iterative processes with 5 or 10 iterations are examined to determine how well the classification model can separate objects. Data partitioning is not as dependable or as simple as the CV method, for example. In our simulations, we randomly select 20 percent of the data to be tested for each simulation run. The rest of the data is used for training purposes[42].

3. Methodology

In this investigation, the original 30 features of the WDBC data were used as input. One feature was decreased with the help of ICA to examine how well the classifier performed when making a choice about breast cancer. Consequently, the model depicted in Figure 3 is used to analyze data from the WDBC, which contains 30 characteristics and 569 patients (patients) who were utilized to train and test the models.

Prior to testing the classifier, the number of dimensions in the data is reduced using the ICA algorithm. The data is then separated into subsamples using a 5/10-CV anda 20 percent partitioning method, respectively. It was utilized in the training and testingof the SVM model. It was necessary to examine the findings of the classifier in order to determine how well it performed. Five-tenths CV and twenty-percent partitioning areused to test and train classifiers by dividing the data into smaller groups, respectively.After they've been trained, test data is utilized to determine how well the classifiers perform in terms of diagnosing things like sensitivity, specificity, accuracy, F-score, Youden's index, DP, and ROC curve, among other things. For the purpose of distinguishing breast cancer from other types of cancer, SVM kernels such as linear, quadratic, and RBF are employed to determine which form of separating hyperplane is the most effective[42]

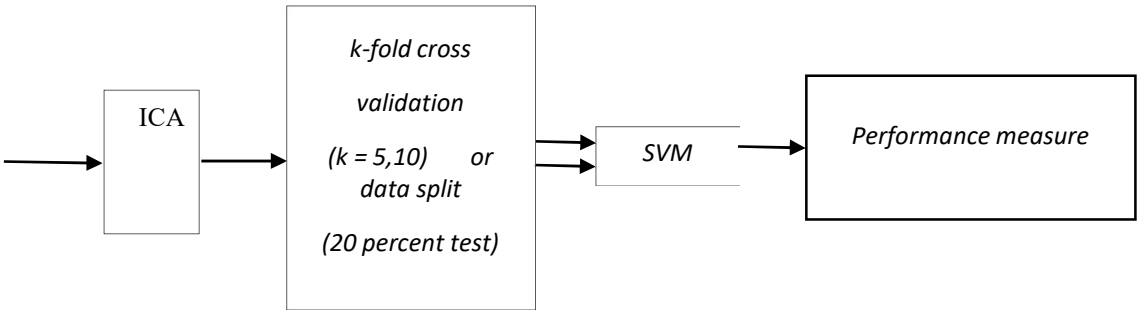


Figure 3 : Research model

4. Results

To train and test the classifiers, we employ a one-dimensional feature vector of WDBCdata reduced by ICA. Using the 5/10 CV approach and 20% of the data as test data, the data's one-dimensionality was determined. The sensitivity value is also used to assess the accuracy of breast cancer classification, as malignant mass classification is more accurate than benign mass classification.

To compute SVM accuracy assessments for linear, polynomial, and RBF kernel functions, kernel function characteristics such as RBF sigma value and polynomial degree were employed.

4.1 Visualizing the data and results

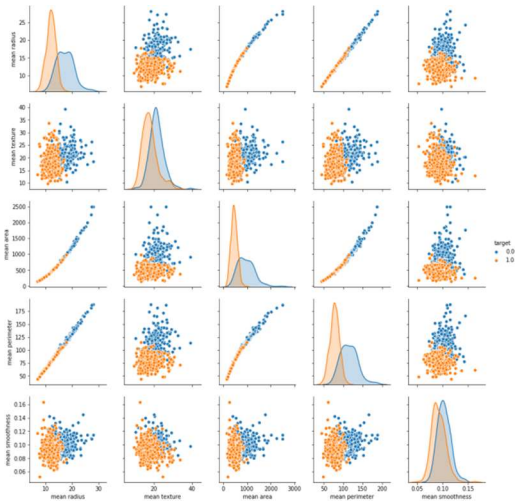


Figure 4 : Data distribution of features/visualization of dataset

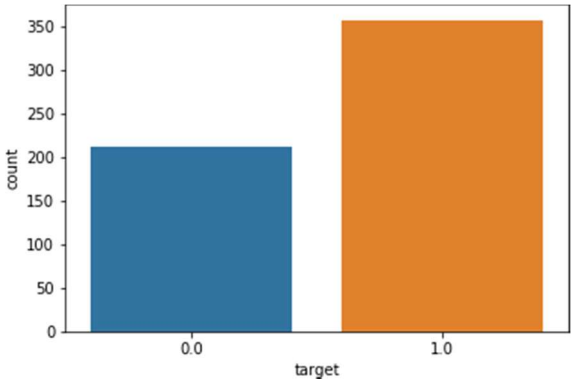


Figure 5 : Comparison of target class after encoding

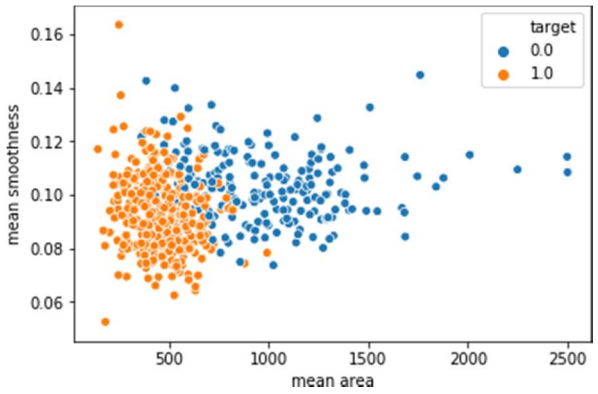


Figure 6 : Classification using mean area and mean smoothness

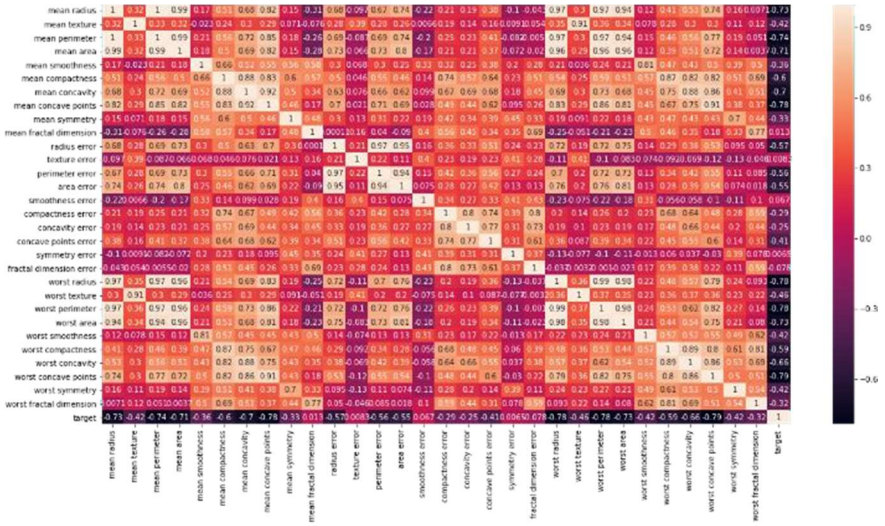


Figure 7 : Correlation using mean area and mean smoothness

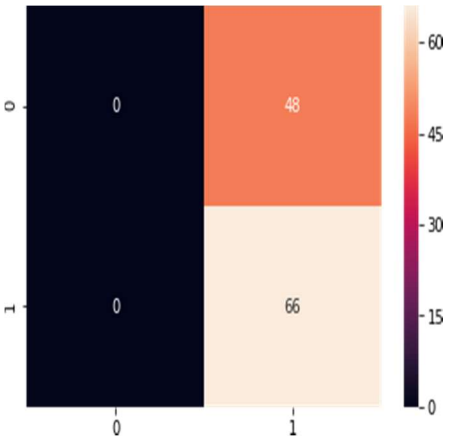


Figure 8 : Confusion matrix

4.2 Improving the model

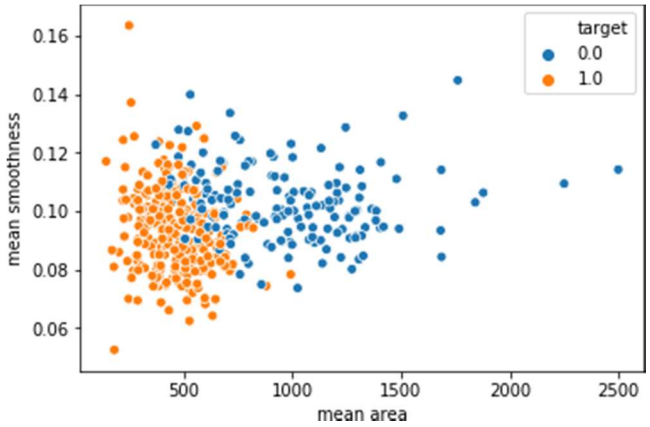


Figure 9 : Improved Classification using mean area and mean smoothness

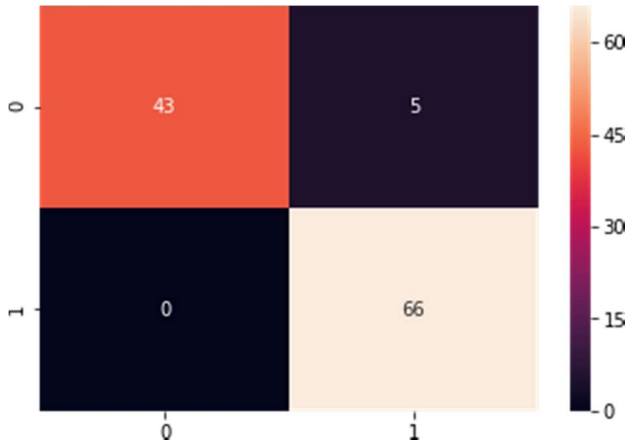
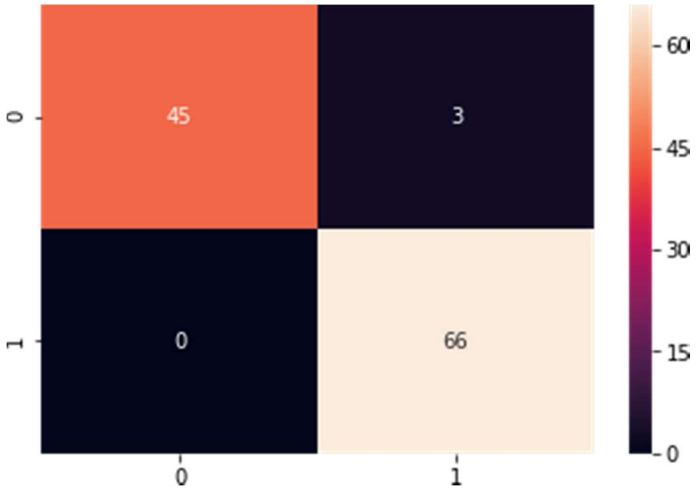


Figure 10 : Confusion matrix

	precision	recall	f1-score	support
0.0	1.00	0.90	0.95	48
1.0	0.93	1.00	0.96	66
avg / total	0.96	0.96	0.96	114

Figure 10 : Improved Confusion matrix



	precision	recall	f1-score	support
0.0	1.00	0.94	0.97	48
1.0	0.96	1.00	0.98	66
avg / total	0.97	0.97	0.97	114

Figure 11 : Improved Confusion matrix

5. Conclusions

This work studies the impact of dimensionality reduction on SVM-based breast cancerdecision support systems using SVMs to classify patients. The original thirty WDBC properties are compared to the reduced one dimension in ICA. The classification accuracy using SVM was 97.47 percent, which is good.

References

[1] I. Christoyianni, E. Dermatas, and G. Kokkinakis, "Fast detection of masses in computer aided mammography," IEEE Signal Processing Magazine, vol. 17, no. 1, pp. 54–64, 2000.

[2] N. Salim, Medical Diagnosis Using Neural Network, Faculty of Information Technology University,2013, http://www.generation5.org/content/2004/MedicalDiagnosis.asp.

[3] A. Tartar, N. Kilic, and A. Akan, "Classification of pulmonary nodules by using hybrid features," Computational and Mathematical Methods in Medicine, vol. 2013, Article ID 148363, 11pages, 2013.

[4] N. Kilic, O. N. Ucan, and O. Osman, "Colonic polyp detection in CT colonography with fuzzy rule-based 3D template matching," Journal of Medical Systems, vol. 33, no. 1, pp. 9–18, 2009.

[5] A.Mert,N.Kilic¸, Anda. Akan, "Evaluation of bagging ensemble method with time-domain feature extraction for diagnosing of arrhythmia beats," Neural Computing and Applications, vol. 24, no. 2, pp. 317–326, 2014.

[6] R. W. Brause, "Medical analysis and diagnosis by neural networks," in Proceedings of the 2nd International Symposium on Medical Data Analysis (ISMDA '01), pp. 1–13, Madrid, Spain,

October 2001.

[7] T. S. Subashini, V. Ramalingam, and S. Palanivel, "Breast mass classification based on cytologicalpatterns using RBFNN and SVM," Expert Systems with Applications, vol. 36,no. 3, pp. 5284–5290, 2009.

[8] M. N. Gurcan, H.-P. Chan, B. Sahiner, L. Hadjiiski, N. Petrick, and M. A. Helvie, "Optimal neural network architecture selection: improvement in computerized detection of microcalcifications," Academic Radiology, vol. 9, no. 4, pp. 420–429, 2002.

[9] A. P. Dhawan, Y. Chitre, C. Bonasso, and K.Wheeler, "Radialbasis-function based classification of mammographic microcalcifications using texture features," in Proceedings of the 17th IEEE Engineering in Medicine and Biology Annual Conference, pp. 535–536, September 1995.

[10] A. T. Azar and S. A. El-Said, "Superior neuro-fuzzy classification systems," Neural Computing and Applications, vol. 23, no. 1, supplement, pp. 55–72, 2012.

[11] M. Jia, C. Zhao, F. Wang, and D. Niu, "A new method for decision on the structure of RBF neural network," in Proceedings of the 2006 International Conference on Computational Intelligence and Security, pp. 147–150,November 2006.

[12] J. K. Sing, S. Thakur, D. K. Basu,M. Nasipuri, and M. Kundu, "High-speed face recognition using self-adaptive radial basis function neural networks," Neural Computing & Applications,

vol. 18, no. 8, pp. 979–990, 2009.

[13] R. Huang, L. Law, and Y. Cheung, "An experimental study: on reducing RBF input dimension by ICA and PCA," in Proceedings of the 2002 International Conference on Machine Learning andCybernetics, vol. 4, pp. 1941–1945, November 2002.

[14] V. N. Vapnik, Statistical Learning Theory, John Wiley & Sons, New York, NY, USA, 1989.

[15] M. F. Akay, "Support vector machines combined with feature selection for breast cancer diagnosis,"Expert Systems with Applications, vol. 36, no. 2, pp. 3240–3247, 2009.

[16] B.Wang, H. Huang, and X.Wang, "A support vector machine based MSM model for financial short-term volatility forecasting," Neural Computing and Applications, vol. 22, no. 1, pp. 21–28, 2013.

[17] M. Pontil and A. Verri, "Support vector machines for 3D object recognition," IEEE Transactions onPattern Analysis and Machine Intelligence, vol. 20, no. 6, pp. 637–646, 1998.

[18] J. Zhou, G. Su, C. Jiang, Y. Deng, and C. Li, "A face and fingerprint identity authentication systembased on multi-route detection," Neurocomputing, vol. 70, no. 4–6, pp. 922–931, 2007.

[19] E. Gumus, N. Kilic, A. Sertbas, and O. N. Ucan, "Evaluation of face recognition techniques usingPCA, wavelets and SVM," Expert Systems with Applications, vol. 37, no. 9, pp. 6404–6408, 2010.

[20] S. Kara, A. Guven, and S. Icer, "Classification of macular and optic nerve disease by principalcomponent analysis," Computers in Biology and Medicine, vol. 37, no. 6, pp. 836–841, 2007.

[21] A. Hyv¨arinen and E. Oja, "Independent component analysis: algorithms and applications," NeuralNetworks, vol. 13, no. 4-5, pp. 411–430, 2000.

[22] M. P. S. Chawla, "A comparative analysis of principal component and independent component techniques for electrocardiograms," Neural Computing and Applications, vol. 18, no. 6, pp. 539–556, 2009.

[23] S. D. Villalba and P. Cunningham, "An evaluation of dimension reduction techniques for one-class classification," Artificial Intelligence Review, vol. 27, no. 4, pp. 273–294, 2007.

[24] W. H.Wolberg,W. N. Street, and O. L. Mangasarian, "Machine learning techniques to diagnose breast cancer from image processed nuclear features of fine needle aspirates," Cancer

Letters, vol. 77, no. 2-3, pp. 163–171, 1994.

[25] K. H. Liu, B. Li, Q. Q. Wu, J. Zhang, J. X. Du, and G. Y. Liu, "Microarray data classification based on ensemble independent component selection," Computers in Biology and Medicine, vol.39, no. 11, pp. 953–960, 2009.

[26] 2013, http://research.ics.tkk.fi/ica/fastica/.

[27] J. Bilski, "The UD RLS algorithm for training feedforward neural networks," International Journal ofApplied Mathematics and Computer Science, vol. 15, pp. 115–123, 2005.

[28] N. Sivri, N. Kilic, and O. N. Ucan, "Estimation of stream temperature in Firtina Creek (Rize-Turkiye)using artificial neural network model," Journal of Environmental Biology, vol. 28, no.1, pp. 67–72, 2007.

[29] O. A. Abdalla, M. H. Zakaria, S. Sulaiman, and W. F.W. Ahmad, "A comparison of feed-forward back-propagation and radial basis artificial neural networks: A Monte Carlo study," in Proceedings of the International Symposium in Information Technology (ITSim '10), vol. 2, pp. 994–998, Kuala Lumpur, Malaysia, June 2010.

[30] B. E. Boser, I. M.Guyon, and V.N.Vapnik, "A training algorithm for optimal margin classifiers," in Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory, pp. 144–152, July 1992.

[31] V.N.Vapnik,TheNature of Statistical Learning Theory, Springer, New York, NY, USA, 1995.

[32] R. Courant and D. Hilbert, Methods of Mathematical Physics, Wiley, New York, NY, USA, 1953.

[33] W. J. Youden, "Index for rating diagnostic tests," Cancer, vol. 3, no. 1, pp. 32–35, 1950.

[34] L. L. Pesce and C. E. Metz, "Reliable and computationally efficient maximum-likelihood estimationof proper binormal ROC curves," Academic Radiology, vol. 14, no. 7, pp. 814–829,

2007.

[35] J. Hamidzadeh, R. Monsefi, and H. S. Yazdi, "DDC: distance-based decision classifier," Neural Computing & Applications, vol.21, no. 7, pp. 1697–1707, 2012.

[36] A. M. Krishnan, S. Banerjee, C. Chakraborty, and A. K. Ray, "Statistical analysis of mammographic features and its classification using support vector machine," Expert Systems with Applications, vol. 37,no. 1, pp. 470–478, 2010.

[37] S. C. Bagui, S. Bagui, K. Pal, and N. R. Pal, "Breast cancer detection using rank nearest neighbor classification rules," Pattern Recognition, vol. 36, no. 1, pp. 25–34, 2003.

[38] N. H. Sweilam, A. A. Tharwat, and N. K. Abdel Moniem, "Support vector machine for diagnosis cancer disease: a comparative study," Egyptian Informatics Journal, vol. 11, no. 2, pp. 81–92, 2010.

[39] O. L. Mangasarian, W. N. Street, and W. H. Wolberg, "Breast cancer diagnosis and prognosis via linear programming," Operations Research, vol. 43, no. 4, pp. 570–577, 1995.

[40] A. Mert, N. Kılıc¸, and A. Akan, “An improved hybrid feature reduction for increased breast cancer diagnostic performance,” Biomedical Engineering Letters, vol. 4, no. 3, pp. 285–291, 2014.

[41] B. Zheng, S. W. Yoon, and S. S. Lam, “Breast cancer diagnosis based on feature extraction usinga hybrid of K-means and support vector machine algorithms,” Expert Systems with Applications, vol. 41, no. 4, pp. 1476–1482, 2014.

[42] Ahmet Mert, Niyazi Kılıç, Erdem Bilgili, Aydin Akan, "Breast Cancer Detection with Reduced Feature Set", Computational and Mathematical Methods in Medicine, vol. 2015, Article ID 265138, 11 pages, 2015. https://doi.org/10.1155/2015/265138.