Institute of Physics Publishing
doi:10.1088/1742-6596/15/1/015

Journal of Physics: Conference Series **15** (2005) 88–94
Sensors & their Applications XIII

Investigation of support vector machine for the detection of architectural distortion in mammographic images

Q Guo[1], J Shao[2] and V Ruiz[1]

[1] Department of Cybernetics, The University of Reading, Reading RG6 6AY, UK
[2] Department of Electronics, University of Kent at Canterbury, Kent CT2 7NT, UK

E-mail: qi.guo@reading.ac.uk, j.shao@kent.ac.uk, v.f. ruiz@reading.ac.uk

Abstract. This paper investigates detection of architectural distortion in mammographic images using support vector machine. Hausdorff dimension is used to characterise the texture feature of mammographic images. Support vector machine, a learning machine based on statistical learning theory, is trained through supervised learning to detect architectural distortion. Compared to the Radial Basis Function neural networks, SVM produced more accurate classification results in distinguishing architectural distortion abnormality from normal breast parenchyma.

1. Introduction

Breast cancer has become the second most common cause of cancer death in women in the UK, after lung cancer. One in every nine women will develop breast cancer at some point in her life, with nearly 40,700 new cases diagnosed each year [1]. Early detection of breast cancer is crucial if treatment is to be successful. Mammography is considered the most effective method for the early detection of breast cancer. Screening programme has been shown to reduce mortality rates by almost half [1].

Computer-aided diagnosis (CAD) techniques and systems are effective in detecting masses and microcalcifications, however, they have been found to fail in the detection of architectural distortion with adequate level of accuracy [2]. The definition of architectural distortion in BI-RADS [3] is as: “The normal architecture of the breast is distorted with no definite mass visible. This includes spiculations radiating from a point and focal retraction or distortion at the edge of the parenchyma.” It has been reported that architectural distortion is the most commonly missed abnormality in false-negative cases [4].

Several studies have been reported on the detection of architectural distortion. Ayres and Rangayyan [5, 6] applied phase portrait maps to characterize oriented texture patterns of the architectural distortion. Mudigonda and Rangayyan [7] studied the use of texture flow-field to detect architectural distortion, based on the local coherence of texture orientation. Matsubara *et al.* [8] used mathematical morphology to detect architectural distortion around the skin line and a concentration index to detect architectural distortion within the mammary gland. There are also a number of studies on the performance of commercial CAD system in the detection of architectural distortion. Burhenne *et al.* [9] obtained a sensitivity of 75% of a commercial CAD system in the detection of architectural distortion. Evans *et al.* [10] reported that a commercial CAD system correctly identified 17 of 20 cases of architectural distortion. However, Baker *et al.* [2] investigated the ability of two commercial CAD systems and found that the sensitivity of these systems to be poor in detecting architectural distortion: fewer than 50% of the cases of architectural distortion were detected. These findings indicate that the detection of architectural distortion is still under developed and there is a need for further research in this area to improve the accuracy of the detection.

In this paper, we present a technique to architectural distortion detection by combing fractal feature extraction with support vector machine classification. The mammographic texture is characterised by Hausdorff dimension [11]. Support vector machine classifier is used to classify ROIs as including architectural distortion or other parenchymal patterns. For comparison purposes, radial basis function neural networks has also been applied to classification. It is found that the SVM outperforms radial basis networks.

2. Dimension Measurement

Hausdorff dimension is the fundamental definition of the fractal dimension in the theory of fractal geometry [11].

Let U be any non-empty subset of n-dimensional Euclidean space R^n. The *diameter* of U is defined as the greatest distance apart of any pairs of points in U, i.e. $|U| = \sup\{|x-y| : x, y \in U\}$. If $\{A_i\}$ is a finite collection of set of diameter at most δ that cover S, i.e. $S \subset \bigcup_{i=1}^{\infty} A_i$ with $0 < |A_i| \le \delta$ for each i, then $\{A_i\}$ is called a $\delta - cover$ of S. Suppose $\hbar$ is a non-negative number. The *$\hbar$-dimensional Hausdorff measure* of a set $\mathrm{S} \subset \mathrm{R}^2$ is defined as

$$\mathcal{H}^{\hbar}(\mathrm{S}) = \lim_{\delta \to 0} \mathcal{H}^{\hbar}_{\delta}(\mathrm{S}) \tag{1}$$

$$\text{with } \mathcal{H}^{\hbar}_{\delta}(\mathrm{S}) = \inf\left\{\sum_i |A_i|^{\hbar} : \{A_i\} \text{ is a } \delta\text{ - }cover \text{ of S}\right\} \tag{2}$$

The *Hausdorff dimension* of S is defined formally as,

$$D_H(\mathrm{S}) = \inf\left\{\hbar : \mathcal{H}^{\hbar}(\mathrm{S}) = 0\right\} = \sup\left\{\hbar : \mathcal{H}^{\hbar}(\mathrm{S}) = \infty\right\} \tag{3}$$

Hausdorff dimension has the advantage of being defined for any set, and is mathematically convenient. However, in general, it is difficult to measure the dimension of a set directly from the definition. Many alternative methods of measuring the dimension of a set have been developed [11]. Regularisation dimension [12] has been proposed as an approximate to the Hausdorff dimension. The advantages of the regularisation dimension are: i) it is more precise than other approximation methods; ii) it is easy to derive an estimator in the presence of noise due to the fully analytical definition.

Let Γ be the graph of a bounded function f: $\mathrm{R} \to \mathrm{R}$ whose support K is a closed bounded ball. Let $\chi(t)$ be a kernel function of Schwartz class S such that: $\int \chi = 1$. Let $\chi_a(t) = \frac{1}{a}\chi(\frac{t}{a})$ be the dilated version of χ at scale a. Let f_a be the convolution of f with χ_a: $f_a = f * \chi_a$ Since $f_a \in S$, the length of its graph Γ on K is finite and given by:

$$L_a = \int_K \sqrt{1 + f_a'(t)^2}\, dt \tag{4}$$

The regularisation dimension of Γ is defined as

$$dim_R(\Gamma) = 1 + \overline{\lim_{a \to 0}} \frac{\log(L_a)}{-\log(a)} \tag{5}$$

3. Classification

3.1. Support Vector Machine

Support vector machine (SVM) is a learning tool based on modern statistical learning theory [13]. It gives some useful bounds on the generalisation capacity of machines for learning tasks. The SVM algorithm constructs a separating hypersurface in the input space. It maps the input space into a higher dimensional feature space through some nonlinear mapping [14].

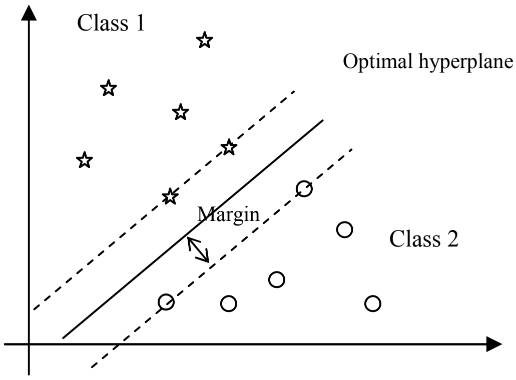


Figure 1. The separation of two classes (indicated by data points marked by "○"s and "☆"s) by an optimal hyperplane.

Let vector $x \in \mathrm{R}^n$ denote a pattern to be classified. Each vector x_i belongs to either of two classes identified by the label $y_i \in \{\pm 1\}$. In addition, let $\{(x_i, y_i)$, i = 1, 2, …, $l\}$ denote a given set of l training examples. The SVM first maps x to a higher dimensional space $\mathcal{H}$ using a nonlinear operator $\Phi(\bullet)$: $\mathrm{R}^n \to \mathcal{H}$. Consider the case when the data is linearly separable in $\mathcal{H}$. The nonlinear SVM classifier is defined as

$$f(x) = w^T \Phi(x) + b \,, w \in \mathcal{H}, b \in \mathrm{R} \tag{6}$$

which is linear in terms of the transformed data $\Phi(x)$, but non-linear in terms of the original data $x \in \mathrm{R}^n$. The SVM classifier is based on the hyperplane that maximises the separating margin between the two classes (Figure 1). When the training set is not separable in $\mathcal{H}$ due to the partial overlapping of the two classes, the previous analysis can be generalised by introducing slack variables ξ_i. The SVM algorithm tries to minimise $\|w\|$ while at the same time separating the data with minimum number of errors. Mathematically, this is done by minimising

$$J(w, \xi) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l} \xi_i \tag{7}$$

with $\xi_i \geq 0$ satisfying the constraint:

$$y_i(w^T \Phi(x_i) + b) \geq 1 - \xi_i \,, i = 1, 2, \ldots, l \tag{8}$$

Here, C is a regularisation parameter controlling the tradeoff between model complexities and training error in order to ensure good generalisation performance.

Using the technique of Lagrange multipliers, one can show that a necessary condition for minimising $J(w, \xi)$ in (7) is that the vector w is formed by a linear combination of the mapped vectors $\Phi(x_i)$, i. e.,

$$w = \sum_{i=1}^{l} \alpha_i y_i \Phi(x_i) \tag{9}$$

where $\alpha_i \geq 0$, i = 1, 2, …, l. are the Lagrange multipliers associated with the constraints in (8). The Lagrange multipliers $\alpha_i \geq 0$, i = 1, 2, …, l, are solved from the dual form of (7), which is

$$\max W(\alpha_1, \alpha_2, ..., \alpha_l) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2} \sum_{i=1}^{l} \sum_{i=1}^{l} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \tag{10}$$

subject to

$$0 \le \alpha_i \le C \text{ , } i = 1, 2, \ldots, l \tag{11}$$

$$\sum_{i=1}^{l} \alpha_i y_i = 0 \tag{12}$$

The kernel function in an SVM plays the central role of implicitly mapping the input vector into a high-dimensional feature space. Typical choices for kernel function are: Gaussian Radial Basis Function (RBF), Polynomial, Sigmoidal, Inverse multiquadratic, etc. The polynomial kernels and Gaussian RBF kernels have been applied in this study. They are defined as follows.

Polynomial kernel: $$K(x, y) = ((x \cdot y) + 1)^p \tag{13}$$

Gaussian RBF kernel: $$K(x, y) = \exp(\frac{-\|x - y\|^2}{\sigma}) \tag{14}$$

where $\sigma \in \mathrm{R}$, the width for the Gaussian RBF function, and $p \in \mathrm{N}$, the degree of the polynomial function.

3.2. Radial Basis Function Neural Networks

Radial Basis function neural network (RBF) [15] is a major class of neural network model, in which the activation of a hidden unit is determined by the distance between the input vector and a prototype vector. RBF neural networks can be viewed as a non-linear mapping between a set of inputs and a set of outputs. Moreover, RBF neural networks can be regarded as linear-in-the-parameters models which have some unique computational advantages over other architectures of neural networks.

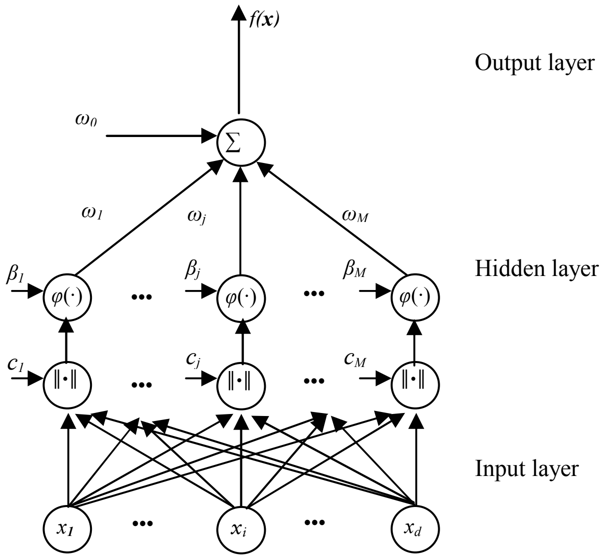


Figure 2. An RBF neural networks with d inputs, M hidden units and one output unit.

An example of RBF neural networks with d inputs, M hidden units and one output unit is shown in Figure 2. The mapping between inputs and outputs of RBF neural networks can be formed as

$$\begin{cases} f_{rbf}(x) = \sum_{t=1}^{M} \omega_j \varphi_j(\alpha_j, \beta_j) + \omega_0 \\ \alpha_j = \left\| x - c_j \right\| \end{cases} \tag{15}$$

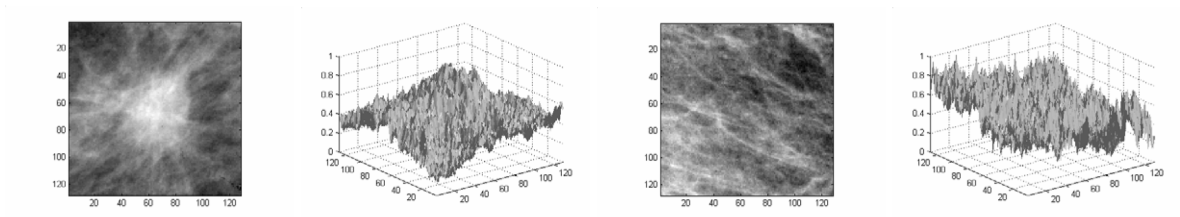
where x = $[x_1, \ldots, x_d]^{\mathrm{T}} \in \mathrm{R}^d$ is input vector, ω_j, $j = 1, \ldots, M$, denotes the weights, ω_0 denotes the bias, $\varphi(\cdot)$, $j = 1, \ldots, M$, are bias functions from R^d to R^1, $\|\bullet\|$ denotes the Euclidean norm, $c_j \in \mathrm{R}^d$, $j = 1, \ldots, M$ are RBF centres, β_j, $j = 1, \ldots, M$ are parameters of basis functions. Typical choices for $\varphi_j(\cdot)$ are: Linear function, Cubic function, Thin-plate-spline function, Gaussian function, Multiquadratic function. Gaussian function has been used in our study, which is defined as

$$\varphi_j(\alpha_j, \beta_j) = \exp\left(-\frac{\alpha_j^2}{\beta_j^2}\right) \tag{16}$$

4. Results and Discussion

In this experiment, we selected a set of 40 Region of interest from MIAS database [16]. Each ROI is of size 128x128 pixels. Spatial resolution is 0.2 mm/pixel. The set includes 19 ROIs with architectural distortion and 21 ROIs with normal tissue patterns. The detection of architectural distortion in this study is considered as a two-class pattern recognition task. The two classes are "architectural distortion" and "normal tissue". The regularisation dimension of each sampled image is computed from equation (4) and equation (5) then input into classifier as a feature.

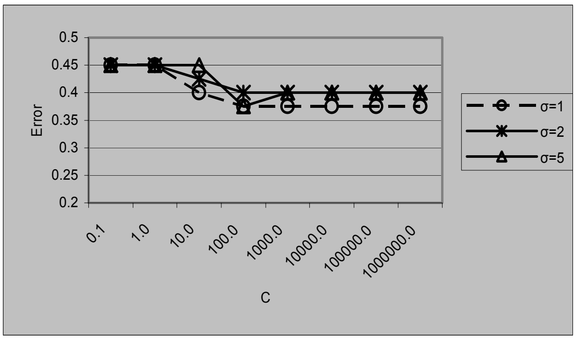
Figure 3 shows two typical images of "architectural distortion" and "normal tissue" and their fractal surface respectively. A clear difference can be observed between the two cases in terms of the surface irregularity. The regularisation dimension appears to hold promise as an effective way of characterising the feature of mammographic images.

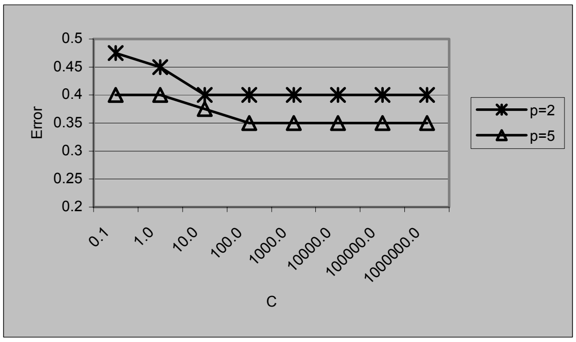


(a) (b) (c) (d)

Figure 3. Sampled images and their surfaces. (a) "architectural distortion"; (b) surface of "architectural distortion" sample; (c) "normal tissue"; (d) surface of "normal tissue".

Generalisation error was regarded as a figure of merit in our evaluations. Generalisation error was defined as the total number of incorrectly classified examples divided by total number of examples classified. Cross-validation has been used in the model selection in our study. The data samples are divided to 4 subsets, each of which has 10 data samples. The classifier is trained 4-times: In the *i*th (i = 1,…, 4) iteration, the classifier is trained on all subsets except the *i*th one. The average of these 4 errors is a rather good estimate of the generalisation error.





(a) (b)

Figure 4. Plot of generalisation error versus regularisation parameter *C*. (a) a Gaussian RBF kernel with width σ=1, 2 and 5, (b) a polynomial kernel with orders p =2 and 5.

In figure 4(a), we summarise the results for the trained SVM when a Gaussian RBF kernel was used. The estimated generalisation error is plotted for different values of the width σ (1, 2, and 5). Similarly, in figure 4(b) we summarise the results when the polynomial kernel was used, the estimated generalisation error is plotted versus the regularisation parameter C for kernel order p = 2 and p = 5. For the Gaussian RBF kernel, we found that the best error level is achieved over a wide range of parameter settings (e.g., when σ = 1 and C is in the range of 10^2-10^6); a better error level was achieved by polynomial kernel when p =5 and C is in the wide range from 10^2-10^6. These results indicate that the performance of the SVM classifier is not very sensitive to the values of the model parameters. The least number of support vectors is 54% when C is set to 100 for polynomial kernel. It is shown that the best classification level using SVM is 72.5% of correct answers.

Table 1 Number of training errors and test errors.

Number of hidden layers	15	16	17	18	19	20
Training	25%	25%	25%	25%	45%	25%
Test	35%	35%	35%	35%	60%	35%

Radial basis function neural networks is also applied for classification. Table 1 illustrates the number of hidden layers chosen has an influence on both the training error and test error. The results show that the error level using RBF is 65% of correct answers. Compared RBF classifier with SVM, results have shown that SVM approach outperforms the RBF neural networks.

5. Conclusion

In this paper, we applied SVM to the detection of architectural distortion in mammographic images by employing the Hausdorff dimension concept. Compared to the radial basis function neural networks, SVM produces better classification results. Future work will be directed toward the finding other relevant features; analysing the complete mammograms for the detection of the architectural distortion; the use of large database of mammograms.

References

[1] Cancer Research UK, *http://www.cancerresearchuk.org /breastcancer/breastcancer*, 2005

[2] Baker J A, Rosen E L, Lo J Y, Gimenez E I, Walsh R and Soo M S 2003 *A. J. Roentgen.* **181** 1083-88

[3] 1998 *Illustrated breast imaging reporting and data system (BI-RADS)*, 3rd ed. (Reston, VA: American College of Radiology)

[4] Burrell H C, Sibbering D M, Wilson A R M, Pinder S E, Evans A J, Yeoman L J, Elston C W, Ellis I O, Blamey R W and Robertson J F R 1996 *Radiol.* **199** 811-817

[5] Ayres F J and Rangayyan R M 2005 *IEEE Eng. in Med. and Biol. Mag.* 59-67

[6] Ayres F J and Rangayyan R M 2003 *Proc. of the 25th Int. Conf. of the IEEE Eng. in Med. and Biol. Society* (Cancun, Mexico) 886-889

[7] Mudigonda N R and Rangayyan R M 2001 *Proc. of Biovision* (Bangalore, India) 76-81

[8] Matsubara T, Ichikawa T, Hara T, Fujita H, Kasai S, Endo T and Iwase T 2003 *Int. Cong. Series: Proc. of the 17th Int. Cong. and Exhib. on Comp. Assisted Radiol. and Surgery.* (London: Elsevier) 950-955

[9] Burhenne L J W, Wood S A, D'Orsi C J, Feig S A, Kopans D B, O'Shaughnessy L F, Sickles E A, Tabar L, Vyborny C J and Castellino R A *Radiol.* **215** 554-562

[10] Evans W P, Burhenne L J W, Laurie L, O'Shaughnessy K F and Castellino R A 2002 *Radiol.* **225** 182-189

[11] Falconer K 1990 *Fractal geometry – mathematical foundations and applications* (NY: Wiley)

[12] Roueff F and Levy Vehel J 1998 *Fractals 98*

[13] Vapnik V 1998 *Statistical learning theory* (NY: Wiley)

[14] Muller K R, Mika S, Ratsch G, Tsuda K and Scholkopf B 2001 *IEEE Trans. Neu. Net.* **12** 181-201

[15] Bishop C M 1995 *Neural networks for pattern recognition* (Oxford University Press)

[16] Suckling J, Dance D R, Lewis D J and Blacker S G 1994 *2nd Int. Workshop on Dig. Mamm.* (Int. Cong. Series 1069) ed A Gale and Astley S M et al (York, England: Excerpta Medica) 315-324