ANALYSIS OF SVM PARAMETRIZATION IN THE CLASSIFICATION OF MAMMOGRAPHIC TEXTURE IMAGES

Álvaro Henrique de Araújo Rungue* Alexei Manso Corrêa Machado† Pedro Augusto Pinho Ferraz* Bernardo Augusto Godinho de Oliveira* Thiago Melo Machado-Coelho‡ Willian Antônio dos Santos*

*Universidade Federal de Minas Gerais
Faculdade de Medicina
Av. Prof. Alfredo Balena 190, Belo Horizonte, MG, Brasil

†Pontifícia Universidade Católica de Minas Gerais
Programa de Pós-Graduação em Engenharia Elétrica
Av. Itaú 525, 30535-012, Belo Horizonte, MG, Brasil

‡Universidade Federal de Minas Gerais
Programa de Pós-Graduação em Engenharia Elétrica
Av. Antônio Carlos 6627, 31270-901, Belo Horizonte, MG, Brasil

Email: alvaro.rungue@sga.pucminas.br, alexei@pucminas.br, pferraz@sga.pucminas.br, bernardo.godinho@sga.pucminas.br, thmmcoelho@ufmg.br, willian.santos.838460@sga.pucminas.br,

Abstract— Studies indicate that breast density is related to the risk of developing cancer since dense breast tissue can hide lesions, causing cancer to be detected at later stages. In this paper we classification method using support vector machines (SVM) associated to data reduction techniques to classify mammographic texture. An analysis of the parameters that influence the effectiveness of texture classification is also provided. Experiments were conducted on a set of 4,000 mammographic exams from which regions of interest representing the most significantly part of the texture of the breast tissue were extracted. Compared to other quantitative results found in the literature, the proposed multi-class SVM method using the radial basis function kernel and tuned parameters proved to be superior while classifying mammographic texture, reaching up to 99% of precision for 10% of recall.

Keywords— Mammography, Image classification, Texture descriptors, Support Vector Machines

Resumo— Estudos indicam que a densidade das mamas está relacionada com o risco de desenvolvimento de câncer, pois mamas com tecidos mais densos podem esconder pequenas lesões e levar sua detecção a estágios mais avançados. Neste artigo, é proposto um método de classificação da densidade das mamas, associado a técnicas de redução de dimensionalidade, para classificar texturas de imagens mamográficas. Os experimentos foram realizados em um conjunto de dados de 4000 exames mamográficos, dos quais as regiões de interesse que representam a parte mais significativa da textura do tecido foram extraídas. O método proposto utilizando um SVM multi-classe e otimização de parâmetros alcançou uma taxa de precisão de 99%, para valores de revocação de 10% das amostras, valores estes superiores aos publicados na literatura.

Palavras-chave— Classificação de Imagens, Descritor de Textura, Mamografias, Máquinas de Vetores de Suporte

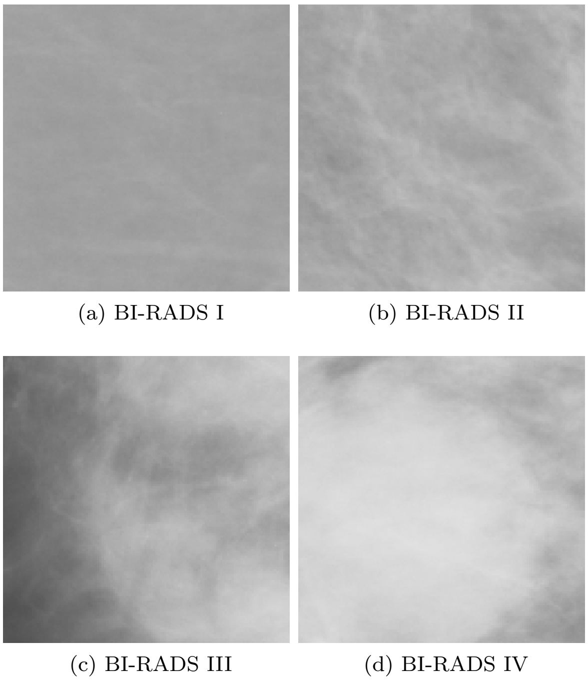
1 Introduction

According to statistics provided by the Brazilian Institute of Cancer (INCA) (Instituto Nacional de Câncer José Alencar Gomes da Silva, 2016), breast cancer was responsible for 17,488 deaths in Brazil, representing 16.8% of cancer mortality in women (World Health Organization, 2014) in 2014. Studies indicate that breast density is related to the risk of developing breast cancer since dense breast tissue can hide lesions, causing cancer to be detected at later stages. Therefore, the analysis of mammographic images is the main screening tool for cancer detection, as it allows for radiologists to evaluate and rate breast density on the basis of visual inspection.

In 1976, the American College of Radiology proposed the Breast Imaging Reporting Data System scale (BI-RADS) to represet 4 levels of breast densities (Wolfe, 1976): BI-RADS I, almost entirely fatty tissue; BI-RADS II, scattered fibroglandular tissue; BI-RADS III, heterogeneously dense tissue; and BI-RADS IV, extremely dense tissue. Fig. 1d shows typical examples of tissue densities classified based on the BI-RADS scale.

Although mammography is still the best way to screen for cancer and to evaluate the BI-RADS levels of density, visual inspection presents some problems related to human error and subjectivity. Furthermore, the breast may be composed of heterogeneous tissues so that the classification becomes more error-prone. Fig. 2 shows patches of mammography exams that were classified into the 4 BI-RADS levels, even though their textures look very similar. In fact, these examples are not exceptions. Databases of BI-RADS texture examples can be very heterogeneous, making clas-

Figure 1: Typical examples of BI-RADS densities.

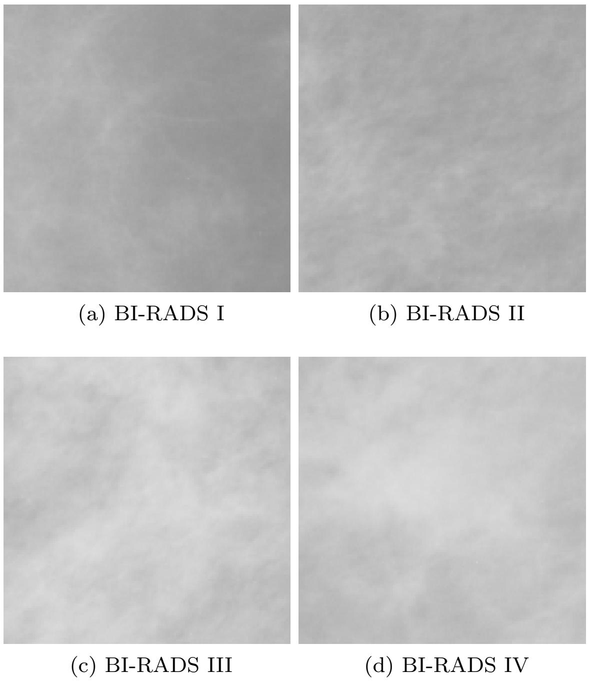


(a) BI-RADS I (b) BI-RADS II

(c) BI-RADS III (d) BI-RADS IV

Source: Aachen University of Technology, IRMA project.

Figure 2: Examples of BI-RADS images with similar texture.



(a) BI-RADS I (b) BI-RADS II

(c) BI-RADS III (d) BI-RADS IV

Fonte: Aachen University of Technology IRMA project

sification a very difficult task for computer-aided diagnosis systems.

Machine learning techniques such as Support Vector Machines (SVM), neural networks and other classifiers based on texture descriptors have been used in the development of systems to support diagnosis. Among all the available methods, SVMs can be considered a good compromise between precision and computational costs. Nevertheless, the results obtained with SVMs are very sensitive and dependent to the correct tunning of their parameters.

This paper presents a comprehensive analysis of the parameters that influence the effectiveness of mammographic texture classification based on SVMs. We also investigate the available topologies to perform multi-class classification using binary SVMs associated to data reduction techniques.

2 Background

Texture description has been frequently addressed in the literature as a data reduction problem, in which a set of descriptors of lower cardinality is pursued. Methods such as the Haralick (Haralick et al., 1973) and Tamura (Tamura et al., 1978) descriptors as well as Principal Component Analysis (PCA) (Maćkiewicz and Ratajczak, 1993) and the Two-dimensional Principal Component Analysis (2DPCA) (Yang et al., 2004) are commonly used to represent texture. The transformed data are then used as input to classifiers like Decision Trees (Safavian and Landgrebe, 1991), SVMs (Vapnik, 2000) and Neural Networks. Convolutional Neural Networks on the other hand will have the whole image as the input and perform implicit description and data reduction.

Studies showing a comparisons between texture descriptors are not uncommon, as the ones presented by (Mohanty et al., 2013) and (Guo et al., 2014), in which features are rated with respect to their ability to discriminate different textures. The work proposed by (Mohanty et al., 2013) applies decision trees in mammographic images to distinguish malign from benign masses. The paper of (Guo et al., 2014) presents a study of statistical and binary texture elements called textons, although their effectiveness to discriminate mammographic densities is unclear.

The management of big data in imaging systems is addressed by (Toews and Wells, 2013) and (Ayma et al., 2015). While the works of (Toews and Wells, 2013) and (Mohanty et al., 2013) propose a web system to support diagnosis, the main goal of (Ayma et al., 2015) is the analysis of distributed aspects of image processing using Hadoop for efficiency gain. Unfortunately, none of them present results for mammographies.

Fuzzy logic has been used to enhance the effectiveness of existing models based on their ability to represent heterogeneous data. This can be seen in the works of (Vieira et al., 2012) and (Hammouche et al., 2015) that present enhanced models using fuzzy logic, fuzzy local binary patterns and fuzzy aura matrices. The fuzzy logic is also used to achieve a better representation of the data in the study of (Li et al., 2015), but applied to modalities other than mammography exams.

Recent studies presented by (Wang et al., 2014) and (Zhang et al., 2015) investigate the use of deep learning techniques as a way to represent visual data and to achieve better discrimination between classes. However, only (Wang et al., 2014) used mammography images to perform the segmentation of masses.

Another important application of texture is in content-based image retrieval systems (CBIR). The work presented by (De Oliveira et al., 2010) proposes a CBIR system to retrieve similar mammographic images from a database, given a query image. A large sample of over 5 thousand images was used in the experiments and the precision was measured based on the amount of retrieved images of the same BI-RADS class. Relevance ranking was obtained from a SVM classification module that used polynomial kernels to enhance discrimination. This work is also one of the few that quantitatively evaluates the results of mammographic texture classification, although the main objective was to estimate the accuracy of the retrieval step, based only on the 10% of the best rated retrieved images. The best results obtained for PCA and 2DPCA were respectively $70,86\%$ and $97,83\%$ of average precision.

More recently, the work presented by (Huang et al., 2017) shows a study on the impact of SVM kernel functions to mammography classification in small and large datasets. Two datasets of $102,294$ samples with 117 features and of 699 samples with 11 features were used in the experiments. The results show that SVM still performs better than other classifiers, although the method has been applied to the classification of masses. Nevertheless, the high precision and sensitivity rates obtained motivates the analysis of SVM parametrization and kernel functions in the context of texture, as will be addresses in the next sections.

3 Texture Characterization

Describing the texture of mammographies is a difficult task because different BI-RADS levels may present similar aspects, as illustrated in Fig. 2. Moreover, to input the images themselves to a classifier may result in loss of efficiency due to the large amount of data. Therefore, data reduction techniques like PCA (Jolliffe, 2002) may be used to extract information from mammographies and transform them into a lower-dimensional representation, improving classification and the required computational time.

PCA is a statistical method that uses an orthogonal transformation to convert a set of possible correlated variables into a set of components that are linearly uncorrelated. The number of principal components should be much lesser than the original number of variables in order for the data reduction to be expressive. The linear transformation of PCA is defined as:

$$Y = BX, \tag{1}$$

where X denotes the original set of n observations with p variables that should be reduced to m variables in set Y. The transformation matrix B is a orthogonal matrix that rotates the original variable space so as to align them with the principal modes of data variance. B can be efficiently computed based on the eigenvectors of the $p \times p$ data covariance matrix.

Differently from PCA, 2DPCA generates vector components instead of scalars. In this case, it is not necessary to transform the images into one-dimensional vectors to calculate the covariance matrix like in PCA. This process results in a much smaller covariance matrix and facilitates the evaluation of the eigenvectors, reducing the computational time required to extract the new features.

The goal of the 2DPCA technique proposed by (Yang et al., 2004) is to project an image A of size $m \times n$ on a vector space X using a linear transformation. The result is a m-dimensional vector Y that is a projected feature vector defined as:

$$Y_{m\times 1} = A_{m\times n} X_{n\times 1}. \tag{2}$$

In order to find a good projection on vector X, the total scatter of the projected samples is used and is defined as the trace of the covariance matrix of the projected feature vectors. So, the following criterion is adopted:

$$J(X) = tr(S_x), \tag{3}$$

where S_x denotes the covariance matrix and $tr(S_x)$ denotes trace of S_x. Maximizing the criterion showed in Equation 3 results in maximizing the total scatter of the resulting projected samples, defined as:

$$G_t = E[(\mathbf{A} - E\mathbf{A})^T(\mathbf{A} - E\mathbf{A})], \tag{4}$$

where G_t denotes the $n\times n$ image covariance (scatter) matrix and A is the images from the dataset. So, alternatively the criterion in Eq. 3 can be defined as follows:

$$J(X)_{1\times 1} = X^T_{1\times n} G_{t(n\times n)} X_{n\times 1}, \tag{5}$$

where X is an unitary column vector.

The unitary vector X that maximizes the criterion defined in Eq. 5 is called the optimal projection axis. The X_{opt} is the vector that maximizes $J(X)$ and could be the eigenvector of G_T

with the higher eigenvalue. In general one value is not enough, so a set of d projection axes, $X_1, \ldots, X_d$, is selected to represent the texture features:

$$Y_k = AX_k \left\{ k = 1, 2, \ldots, d, \right. \tag{6}$$

where Y_k is a family of projected features vectors, $Y_1, Y_2, \ldots, Y_d$ that is called the principal component vectors of image A.

4 Support Vector Machine (SVM)

The support vector machine is a classifier for supervised learning based on statistical learning theory proposed by (Vapnik, 2000). The SVM is essentially a binary classifier, but several extensions were added to allow for multi-class classification problems (Weston, 2000).

The learning mechanism of the SVM consists in determining a hyperplane defined as:

$$f(x) = w.x + b, \tag{7}$$

where w is a vector normal to the hyperplane and $w.x$ denotes the inner product between w and x. Assuming the existence of two classes and sets of points for each class inside the feature space, the SVM determines an hyperplane that splits the feature space so as to maximize the distance between each class and the hyperplane. The subset of points that are used to define this hyperplane are called support vectors.

There are some problems such as the BI-RADs classification in which the feature space cannot be effectively split by a linear function. In this case, a kernel function can be used to map the original feature space into a higher dimensional one where the samples of the two classes are better separated. The most common kernel functions are:

1. linear $\rightarrow K(X_i, X_j) = X_i^T X_j$
2. polinomial $\rightarrow K(X_i, X_j) = (1 + X_i^T X_j)^p$
3. gaussian $\rightarrow K(X_i, X_j) = \exp(-\frac{\|X_i - X_j\|^2}{2\sigma^2})$
4. sigmoid $\rightarrow K(X_i, X_j) = \tanh(\beta_0 X_i^T X_j + \beta_1)$
5. RBF $\rightarrow K(X_i, X_j) = e^{-\gamma ||X_i - X_j||^2}, \gamma > 0$

In cases where there are more than two classes, the SVM should be adjusted to deal with multiclass classification. Three approaches may be used in this case:

1. In the one-against-one (or voting scheme) approach, the k classes are combined two by two resulting in $(k2 - k)/2$ SVMs. Each SVM is trained separately and the new subject to be classified is input to all them. The class that is chosen by the majority of the SVMs is assigned to the subject.

2. In the one-against-all approach, k SVMs are trained, each one for a class, grouping the remaining classes. The class of the associated SVM that reports the largest classification probability is assigned to the new subject.

3. The third way to deal with multiclass classification is to construct a decision tree where the root should have one of the classes against the others and the sub-trees are recursively built with the remaining classes. In this case, $k!$ trees are possible, resulting in a cascade of $k - 1$ SVMs. In the test step, the new subject is input to the SVM at the root of the tree. If the associated class presents a classification probability greater than 0.5 the process stops. Otherwise, the subject is input to the next SVM and the process proceeds until a single class is determined.

5 Methods and Materials

The dataset used in the experiments consists in 4,000 mammographic images gently provided by the IRMA project of the Aachen University of Technology. The size of the images varies from 1024×300 pixels to 1024×800 pixels. The images are equally distributed among all classes and each BI-RADS density is represented by 1,000 images, containing both cranio-caudal (CC) and medio-lateral (MLO) views. The original mammographic exams were previously cropped to extract an region of interest (ROI) of 128×128 pixels and stored in the Portable Network Graphics (PNG) format. The extracted ROIs represent the most significantly part of the texture of the breast tissue, excluding artifacts such as annotation and exam labels from mammographies.

The methods were implemented in Java programming language. The *OpenCV* library was used to perform PCA and the 2DPCA was implemented according to the model described in the previous section. The SVM implementation used the *libsvm* library (Chang and Lin, 2011). The library documentation also provides a guide and highlights the importance of correct parametrization (Hsu et al., 2003).

The number of principal components extracted from the images were 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15 and 20. These values were chosen so as to allow for comparison with the work of (De Oliveira et al., 2010) that used the same dataset and is one of the few to present quantitative results on the classification of mammographic texture. A difference between the two studies is that (De Oliveira et al., 2010) aimed at proposing an image retrieval system instead of a classifier. In order to evaluate the results, the precision of retrieval was measured for the 10% of the retrieved images.

After extracting the principal components, the SVM classification process was performed using a random k-fold cross-validation design, with $k = 100$. In order to achieve better classification rates, two types of training schemes for multiclass classification were investigated: the one-against-one and the decision tree approaches. The decision tree that provided the best overall precision rate had the following structure:

1. The first step used a SVM with BI-RADS I against all the other BI-RADS. In the case the machine returns the class I as the winner, the procedure stops and the subject is assigned to this class, otherwise it proceeds to the next step.

2. The second step uses a SVM with BI-RADS II against BI-RADS III and IV. In the case the machine returns the class II as the winner, the procedure stops and the subject is assigned to this class, otherwise it proceeds to the last step.

3. The last step is a binary SVM between BI-RADS III and IV.

Five kernel functions were investigated: linear, polynomial, Gaussian, sigmoid and radial basis function, as described in section 3. The parameter estimation was performed for each configuration through the method of grid searching, in which the values of all parameters are jointly varied according to a specific function, not necessarily linear (Hsu et al., 2003).

6 Results

The experiments revealed that the PCA was more effective using a sequence of SVMs just like a decision tree and the nodes are a SVM classifier, for all numbers of components. The radial basis function (RBF) also showed to be the best kernel, although the linear and sigmoid functions presented similar results. As suggested by (Hsu et al., 2003), the RBF kernel was therefore preferred since it has only two parameters to estimate, which makes the grid search more computationally efficient. Moreover, it is more stable than some cases of polynomial kernel that may go to infinite and is a generalization of the sigmoid kernel.

The best precision rate (98.00%) was obtained with 15 components. Table 1 presents the results achieved by the proposed method that can be compared to the ones achieved by (De Oliveira et al., 2010) (MammoSYS), used as the baseline for comparison. The best results of MammoSYS were achieved using 20 components (70.86%) at 10% of recall. The results of the proposed classifier design was significantly superior for all number of components.

Table 1: Results of Classification using PCA for different numbers of components m, at 10% and 100% of recall, compared to the results of MammoSYS.

m	**MammoSYS**	**Class. at 10%**	**Class. at 100%**
1	67.86%	70,80%	41,55%
2	68.30%	73,00%	49,82%
3	69.05%	83,30%	57,17%
4	69.56%	91,00%	62,80%
5	69.98%	89,20%	64,12%
6	69.8%	92,70%	64,57%
7	70.39%	93,80%	64,10%
8	70.27%	93,50%	64,17%
9	70.24%	97,00%	63,20%
10	70.45%	96,00%	62,90%
15	70.44%	98,00%	60,57%
20	70.86%	97,30%	60,60%

Experiments using 2DPCA were more effective using the one-against-one design. The radial basis function (RBF) also showed to be the best kernel, yielding a precision rate of 99.80% with 5 components. Table 2 presents the results achieved by the proposed method that can be compared to the ones achieved by MammoSYS, whose best results were also achieved using 5 components (97.83%). Fig. 3 shows the average precision-recall plot, comparing the results between PCA and 2DPCA for the classification of mammographic texture.

Table 2: Results of Classification using 2DPCA for different numbers of components d. at 10% and 100% of recall, compared to the results of MammoSYS.

d	**MammoSYS**	**Class. at 10%**	**Class. at 100%**
1	83,86%	88,70%	58,52%
2	86,03%	94.80%	63,38%
3	87,96%	98,00%	65,52%
4	90,87%	99,50%	67,72%
5	97,83%	99,80%	68,41%
6	97,67%	99,50%	68,91%
7	97,09%	99,80%	69,08%
8	97,00%	99,10%	68,80%
9	96,46%	98,90%	68,61%
10	96,22%	96,00%	68,36%
15	95,50%	98,30%	66,69%
20	93,85%	85,00%	65,16%

The results presented in Tables 1 and 2 corroborate the importance of parameter estimation for SVMs. By tunning the kernel function, the results obtained with PCA become comparable to the ones achieved by 2DPCA on MammoSYS. Fig. 4 illustrates how the results of PCA can be influenced by the choice of the parameters to achieve results similar to 2DPCA at MammoSYS.

Table 3 summarizes the best SVM parameters a grid search was performed starting with randomly values of cost and gamma. The achieved

Figure 3: Precision × recall curve for the average precision using 2DPCA with 6 components and PCA with 15 components.

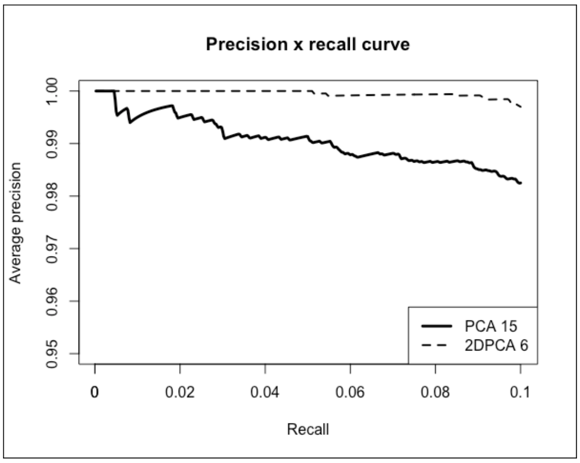
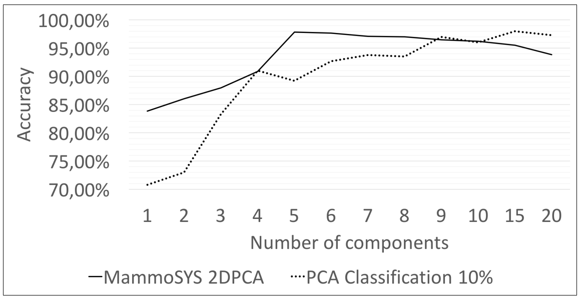


Figure 4: A comparison between the precision of MammoSYS using 2DPCA and the proposed method using PCA for different number of components.



result are shown in Table 3.

7 Conclusion

In this paper we presented an analysis of SVM parametrization applied to the problem of mammographic texture classification. Linear models of data reduction like PCA and 2DPCA were used to extract the features from breast images. Different designs for multi-class classification were also investigated, such as combining multiples SVMs in a decision tree rather than using conventional scoring systems. Compared to other quantitative results found in the literature, the proposed multi-class SVM using the radial basis function kernel with tuned parameters proved to be superior while classifying mammographic texture, reaching up to 99.8% of precision for 10% of recall.

Much improvement can still be achieved in the classification of the whole dataset. Mammographic texture can be very heterogeneous, causing the classifier to fail in around 30% of the cases. Methods such as the convolutional neural networks has not yet been applied to the classification of BI-RADS density levels and are a field for future work. Additionally, other descriptors can be used or combined with the principal components to improve accuracy.

Table 3: SVM best parameters

	PCA		2DPCA	
	Cost	Gamma	Cost	Gamma
1	0.5	0.5	512	0.03125
2	32	8	128	0.03125
3	128	2	2	0.5
4	8	8	2	0.125
5	8	8	2	0.125
6	32	2	2	0.125
7	2048	5	2	0.03125
8	8192	0.125	2	0.03125
9	2048	0.125	2	0.03125
10	2048	0.125	2	0.03125
15	2	2	2	0.03125
20	2	2	0.5	0.0078125

Acknowledgment

The dataset used in this work is a courtesy of TM Deserno, Dept. of Medical Informatics, RWTH Aachen, Germany.

References

Ayma, V., Ferreira, R., Happ, P., Oliveira, D., Feitosa, R., Costa, G., Plaza, A. and Gamba, P. (2015). Classification algorithms for big data analysis, a map reduce approach, *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* **40**(3): 17.

Chang, C.-C. and Lin, C.-J. (2011). LIBSVM: A library for support vector machines, *ACM Transactions on Intelligent Systems and Technology* **2**: 27:1–27:27. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

De Oliveira, J. E., Machado, A. M., Chavez, G. C., Lopes, A. P. B., Deserno, T. M. and Araújo, A. d. A. (2010). Mammosys: A content-based image retrieval system using breast density patterns, *Computer methods and programs in biomedicine* **99**(3): 289–297.

Guo, Z., Zhang, Z., Li, X., Li, Q. and You, J. (2014). Texture classification by texton: statistical versus binary, *PloS one* **9**(2): e88073.

Hammouche, K., Losson, O. and Macaire, L. (2015). Fuzzy aura matrices for texture classification, *Pattern Recognition* **53**: 212–228.

Haralick, R. M., Shanmugam, K. and Dinstein, I. H. (1973). Textural features for image classification, *Systems, Man and Cybernetics, IEEE Transactions on* **SMC-3**(6): 610–621.

Hsu, C.-W., Chang, C.-C., Lin, C.-J. et al. (2003). A practical guide to support vector classification.

Huang, M.-W., Chen, C.-W., Lin, W.-C., Ke, S.-W. and Tsai, C.-F. (2017). Svm and svm ensembles in breast cancer prediction, *PloS one* **12**(1): e0161501.

Instituto Nacional de Câncer José Alencar Gomes da Silva (2016). Síntese de resultados e comentários, `http://www.inca.gov.br/estimativa/2016/sintese-de-resultados-comentarios.asp`. Accessed 23-05-2016.

Jolliffe, I. (2002). *Principal component analysis*, Wiley Online Library.

Li, Y., Huang, Q., Xie, W. and Li, X. (2015). A novel visual codebook model based on fuzzy geometry for large-scale image classification, *Pattern Recognition* **48**(10): 3125–3134.

Maćkiewicz, A. and Ratajczak, W. (1993). Principal components analysis (pca), *Computers & Geosciences* **19**(3): 303–342.

Mohanty, A. K., Senapati, M. R., Beberta, S. and Lenka, S. K. (2013). Texture-based features for classification of mammograms using decision tree, *Neural Computing and Applications* **23**(3-4): 1011–1017.

Safavian, S. R. and Landgrebe, D. (1991). A survey of decision tree classifier methodology, *IEEE transactions on systems, man, and cybernetics* **21**(3): 660–674.

Tamura, H., Mori, S. and Yamawaki, T. (1978). Textural features corresponding to visual perception, *Systems, Man and Cybernetics, IEEE Transactions on* **8**(6): 460–473.

Toews, M. and Wells, W. M. (2013). Efficient and robust model-to-image alignment using 3d scale-invariant features, *Medical image analysis* **17**(3): 271–282.

Vapnik, V. N. (2000). The nature of statistical learning theory, ser. statistics for engineering and information science, *New York: Springer* **21**: 1003–1008.

Vieira, R. T., Ferraz, C. T., de Melo Langoni, V. and Gonzaga, A. (2012). Uma nova proposta para análise de textura local por conjuntos fuzzy, *Anais do VIII Workshop de Visao Computacional*.

Wang, Z., Yu, G., Kang, Y., Zhao, Y. and Qu, Q. (2014). Breast tumor detection in digital mammography based on extreme learning machine, *Neurocomputing* **128**: 175–184.

Weston, J. A. E. (2000). *Extensions to the support vector method*, PhD thesis, Citeseer.

Wolfe, J. N. (1976). Breast patterns as an index of risk for developing breast cancer, *American Journal of Roentgenology* **126**(6): 1130–1137.

World Health Organization (2014). Brazil cancer profile, `http://www.who.int/entity/cancer/country-profiles/bra_en.pdf`. Accessed 23-05-2016.

Yang, J., Zhang, D., Frangi, A. F. and Yang, J.-y. (2004). Two-dimensional pca: a new approach to appearance-based face representation and recognition, *IEEE transactions on pattern analysis and machine intelligence* **26**(1): 131–137.

Zhang, Q., Yang, L. T. and Chen, Z. (2015). Deep computation model for unsupervised feature learning on big data, *IEEE Transactions on Services Computing* **9**(1): 161–171.