Comparative performance analysis of machine learning classifiers and dimensionality reduction algorithms in detection of childhood pneumonia

Rafael T. Sousa[1], Oge Marques[2], Iwens I. G. Sene Jr[1]
Anderson Soares[1], Leandro L. G. de Oliveira[1]

[1]Instituto de Informática – Universidade Federal de Goiás (UFG)
Caixa Postal 131 - 74.001-970 – Goiânia – GO – Brazil

[2]Department of Computer & Electrical Engineering and Computer Science
Florida Atlantic University (FAU) – Boca Raton – USA

`rafaelsousa@inf.ufg.br`

Abstract. *This work extends* PneumoCAD*, a Computer-Aided Diagnosis system for detecting pneumonia in infants using radiographic images[Oliveira et al. 2008], with the aim of improving the system's accuracy and robustness. We implement and compare three contemporary machine learning classifiers, namely: Naïve Bayes, K-Nearest Neighbor (KNN), and Support Vector Machines (SVM), combined with three dimensionality reduction algorithms: Sequential Forward Elimination (SFE), Principal Component Analysis (PCA), and Kernel Principal Component Analysis (KPCA) . Results of our experiments demonstrate that the Naïve Bayes classifier combined with KPCA produces the best overall results.*

Resumo. *Este trabalho complementa o* PneumoCAD*, um sistema de auxílio a diagnóstico para detecção de pneumonia infantil usando imagens radiográficas [Oliveira et al. 2008], com o objetivo de aprimorar a acurácia e robustez do sistema. Nós implementamos e comparamos três classificadores conteporâneos, que são: Naïve Bayes, K-Nearest Neighbor (KNN), e Support Vector Machines (SVM), combinados com três algoritmos de redução de dimensionalidade: Sequential Forward Elimination (SFE), Principal Component Analysis (PCA), e Kernel Principal Component Analysis (KPCA). Os resultados demonstram que o Naïve Bayes combinado com o KPCA produz os melhores resultados.*

1. Introduction

Pneumonia is an epidemic disease characterized by acute lower respiratory infection, usually caused by viruses or bacteria and, less commonly, other microorganisms. According to the World Health Organization (WHO), pneumonia is the leading cause of death in children worldwide, killing an estimated 1.2 million children under five years old every year. This number is higher than the mortality rate for several other diseases, such as AIDS, malaria and tuberculosis, combined [WHO 2012].

Currently the best and most widely accepted imaging modality for detecting pneumonia is chest radiographs [WHO 2001]. However, some studies have shown that errors are common in the interpretation of chest radiographs, due to inter-observer variation

[Young and Marrie 1994]. This limitation of human expert-based diagnosis has provided a strong motivation for the use of computer technology to improve the speed and accuracy of the detection process.

A Computer-Aided Diagnosis (CAD) software can be defined as a second opinion in a diagnostic [Doi et al. 1999]. This kind of software is utilized to improve diagnostic accuracy, not as a means of replacing the specialist, but instead working like a second one, which is invariant to many factors that can affect the radiologist's diagnosis, such as eyestrain, distraction, stress and others.

Our approach to the design and implementation of a CAD system, *PneumoCAD*, consists in mimicking the specialist's vision and perception. This method can be summarized in two main steps: first, the eyes see pictures and find out which aspects of those pictures can be used to describe the patient's clinical condition; soon after, the visual cortex makes decisions based on information obtained by the eyes. In computer vision, these steps are called *feature extraction* and *classification by supervised learning*, respectively.

In this work we use the features and dataset employed in previous studies [Oliveira et al. 2008] [Macedo and Oliveira 2012], which have resulted in a full CAD system for pneumonia detection called *PneumoCAD*, which has been applied to assist in diagnostics, as well as to train and improve radiologists' expertise in childhood pneumonia detection using chest radiographs [Macedo and Oliveira 2012]. *PneumoCAD* is currently in prototype stage. Figure 1 shows the prototype diagnosis screen, with the functionality to upload (and annotate) radiographs which show pneumonia. The ultimate goal behind *PneumoCAD* is to create a website that will provide remote diagnosis functionality by analyzing uploaded chest radiographs and processing them using image processing and machine learning algorithms.

This work was geared towards performing a comparative performance analysis of state-of-art classifiers combined with features selection algorithms, to improve *PneumoCAD* accuracy and find out the best classifier for childhood pneumonia detection.

1.1. Related Works

Depeursinge et al.[Depeursinge et al. 2012], compare five different classifiers in the context of multiclass classification of six types of lung tissues, using high-resolution computerized tomography. The feature space is composed by 39 texture features extracted using quincunx wavelet frame coefficients. A simple grid search for best classifier parameters is performed. The SVM classifier achieves the best trade-off between error rate on the training set and generalization, producing good results, which can be further optimized with a feature selection algorithm.

Yao et al. [Yao et al. 2011] also developed a computer-assisted detection system for identifying and measuring pulmonary abnormalities in cases of infection such as H1N1 influenza. Forty chest computerized tomographic examinations were studied using texture analysis and support vector machine to differentiate normal from abnormal lung. The SVM-based approach achieves good results in successfully distinguishing between areas of abnormality, which demonstrates the efficiency of SVM-based approaches to texture classification.

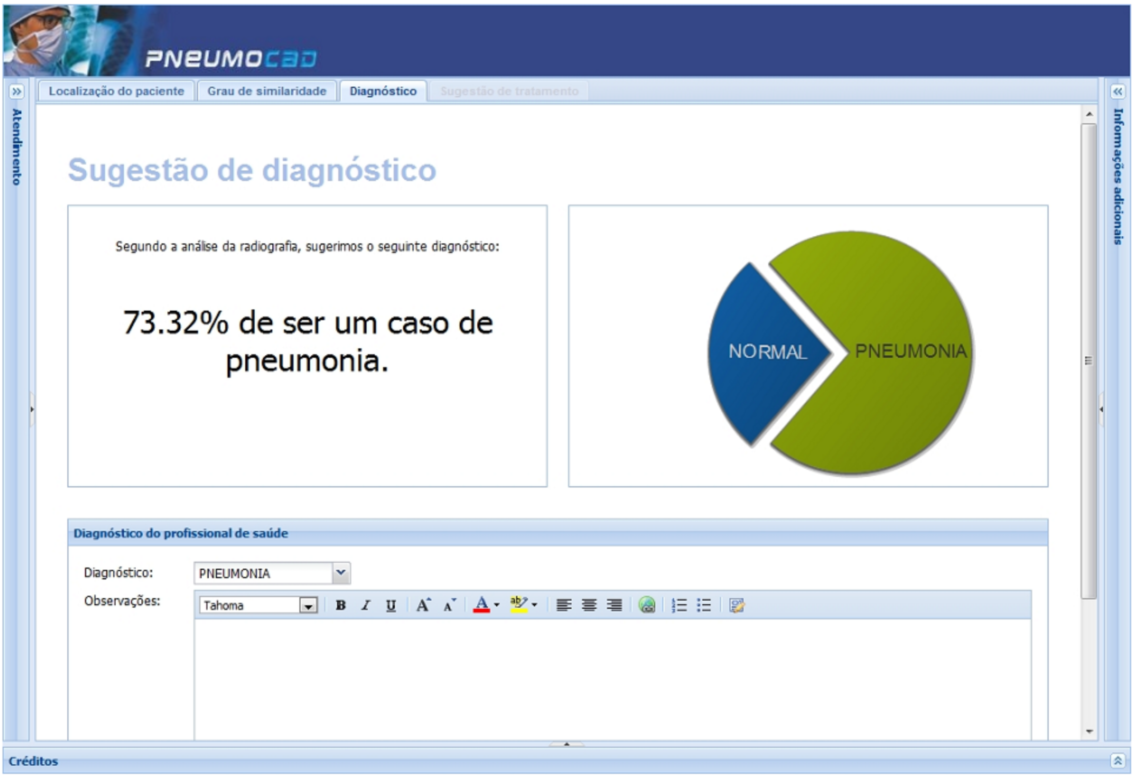

Figura 1. PneumoCAD diagnosis by uploaded radiograph.

1.2. Selected Classifiers

In this paper we apply three different classifiers, in order to find the one that best improves the accuracy of diagnosis with radiographs in *PneumoCAD*, namely: The k-nearest neighbor classifier (kNN), which was used originally in PneumoCAD, Naïve Bayes probabilistic classifier, and non-linear Support Vector Machine (SVM).

The kNN classifier basically responds to inputs as belonging to the class with which it has k nearest neighbors, using some type of metric as the Euclidean distance [Cover and Hart 1967]. The method is easy to apply and computationally fast, but is very sensitive to the curse of dimensionality and the choice of k, which can result in misclassified outliers.

The Naïve Bayes is based on a conditional probability model, which defines the classification by the posterior probability of an entry to belong to one of the known classes. This probability $P(c_i|\vec{v})$ from class c_i, given a feature vector $\vec{v}$, is determined utilizing the Bayes theorem:

$$P(c_i|\vec{v}) = \frac{P(\vec{v}|c_i)P(c_i)}{P(\vec{v})} \tag{1}$$

This method generally has good performance both in computational performance, as well as robustness in classification, including handling of outliers.

SVM maps the input feature vector in a space with higher dimensionality using a kernel function$K(\vec{v_i}, \vec{v_j}) = \langle \phi(\vec{v_i}), \phi(\vec{v_j}) \rangle$, for example, the Gaussian kernel, which is defined as:

$$K(\vec{v_i}, \vec{v_j}) = e^{\frac{-\left\|\vec{v_i}-\vec{v_j}\right\|^2}{2\sigma}} \tag{2}$$

where σ is the width of the Gaussian. In the transformed space, a separating hyperplane is created which separates the positive from the negative examples. Two parallel hyperplanes are created on each side of the first one, with the goal of maximizing the distance between these two hyperplanes, called *margin* [Burges 1998].

Unlike other classifiers, SVM does not have the objective of maximizing performance to the training set, but it is geared towards training generalization, avoiding overfitting and allowing the use of limited data for training [Abe 2010].

2. Methods

The images dataset used in our CAD system consists of 156 8-bit grayscale images obtained with a digital camera, that captured the chest X-rays images at a resolution of 1024×768 pixels. Out of these images, 78 show pneumonia while the remaining 78 do not. Figure **??** shows examples of the images in the dataset. These images were analyzed by two trained radiologists according to WHO guidelines [Levine et al. 1999] [Cherian et al. 2005] which produced the ground truth needed to test the machine learning classifiers used in this work. The radiologists diagnosis was only considered as valid when they agree among themselves.

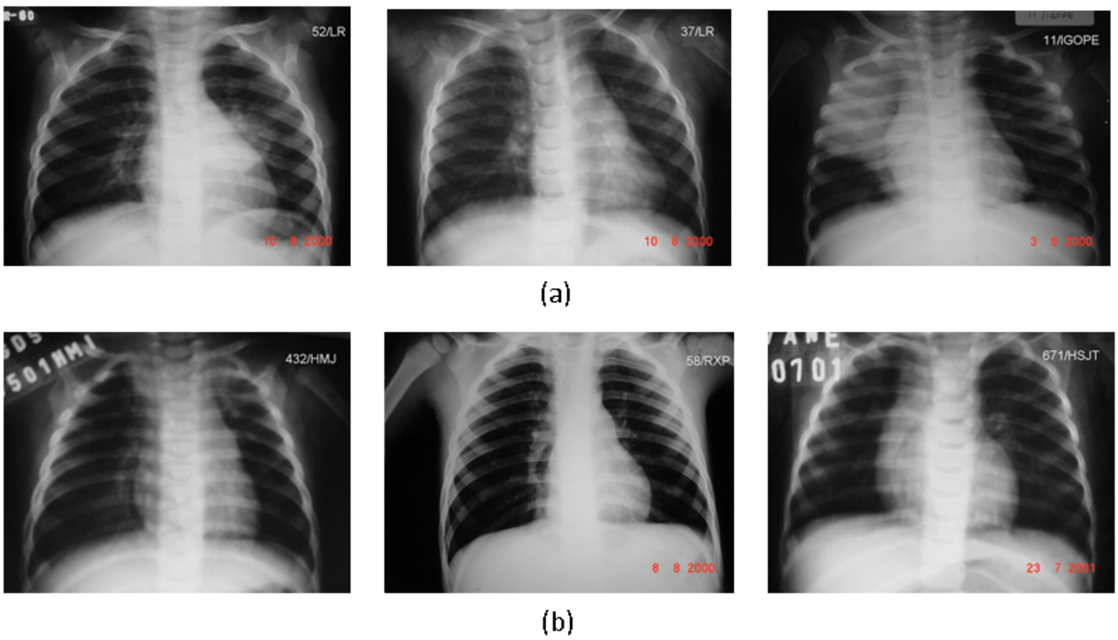

Figura 2. (a) Three chest radiographs positive for pneumonia. (b) Three healthy children radiographs.

All features tested are based on texture: coefficient of variation, contrast, correlation, energy, average energy, entropy, average deviation, difference variance, difference entropy, inverse difference moment, residual mean, sum average, sum entropy, sum variance, suavity, variance, standard deviation [Haralick et al. 1973] [Huang and Dai 2004] [Kokare et al. 2005] [Cheng 2003] [Bashar et al. 2003] [Kokare et al. 2004]. All these features have been extracted in nine subspaces of Haar wavelet.

All tests was made with Matlab along with their basic Toolboxes, Matlab Toolbox for Dimensionality Reduction [van der Maaten 2013], and Weka [Hall et al. 2009].

First was removed all outliers which are out of the interval $\bar{x} - \sigma \leq x \leq \bar{x} + \sigma$, where x is a sample, $\bar{x}$ the feature mean and σ the standard deviation.

We then performed a 10-fold cross-validation test with each classifier. Those who have parameters to be adjusted, were calibrated with a exhaustive search, testing many possible values for each parameter, according to Table 1.

Tabela 1. Parameters interval.

Classifier	Parameter	Variation	
KNN	k	[0; 100]	lin(2)
SVM	C, σ	[1; 30],[1; 30]	lin(1), lin(1)
Naïve Bayes	-	-	-

2.1. Dimensionality Reduction

Based on previous tests with whole feature vector, which result in a insufficiently method (70% correct rate with KNN), we decided to improve our results performing a dimensionality reduction, removing redundant and insignificant features for classification.

All classifiers was tested with each of dimensionality reduction algorithm, which are: Sequential Forward Selection (SFS), which is a simple greedy search algorithm to find the best feature set for each classifier [Ladha and Deepa 2011] [Guyon and Elisseeff 2003]. Principal component analysis (PCA), who maps the data into a new and dimensionally smaller feature space [Hotelling 1933] [van der Maaten et al. 2009], and the Kernel PCA (KPCA), a reformulation of traditional PCA, which realize the mapping using a kernel function [Schölkopf et al. 1998]. The tests was made with a Gaussian kernel.

The SFE algorithm automatically select his new feature space size, but the PCA methods, which generate a new feature space, need previously know how many dimension will have the new space. Based in some observations and tests using different values, we decided to use 13 dimensions.

2.2. Classifiers Evaluation

All tests made was evaluated with Accuracy (correct rate) to compare the overall results and AUC (Area Under Curve ROC calculated by trapezoidal approximation), which has been shown as a better measure to evaluate machine learning classifiers [Huang and Ling 2005].

3. Results

Accuracy results of each classifier with all three dimensionality reduction algorithms, SFE, PCA and KPCA is shows in Figure 4.

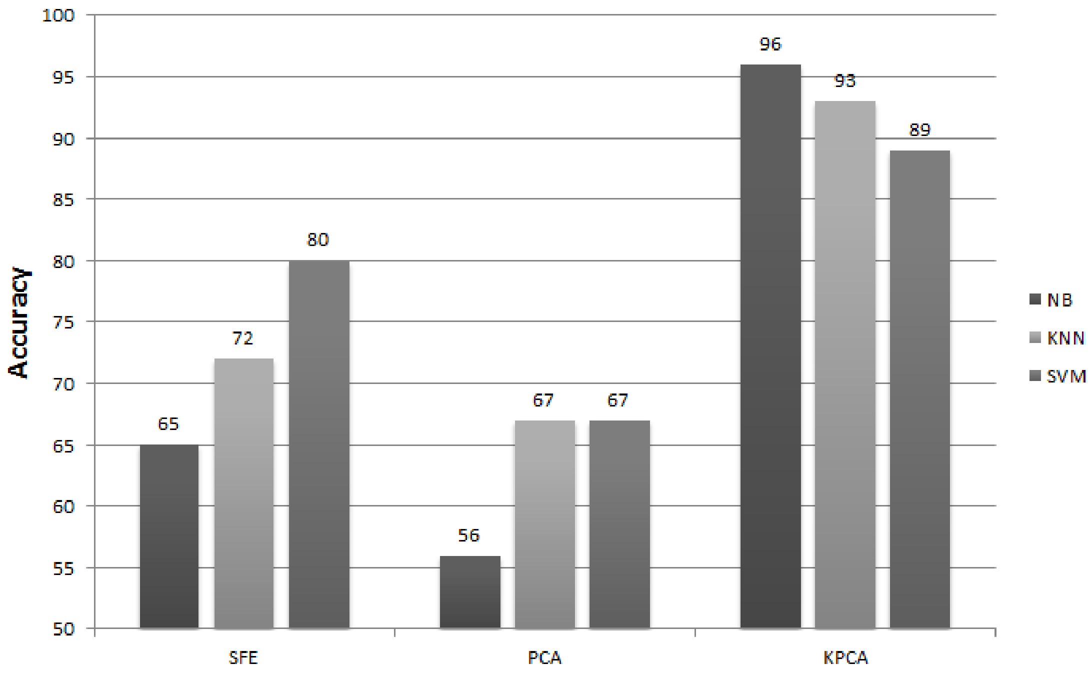

Figura 3. Classifiers accuracy

Following Figure 5,6 and 7 shown the ROC (Receiver Operating Characteristic) curve of each combination and Table 2, 3 and 4 expose the data, as Sensitivity, Specificity and AUC (Area under curve) from same combinations.

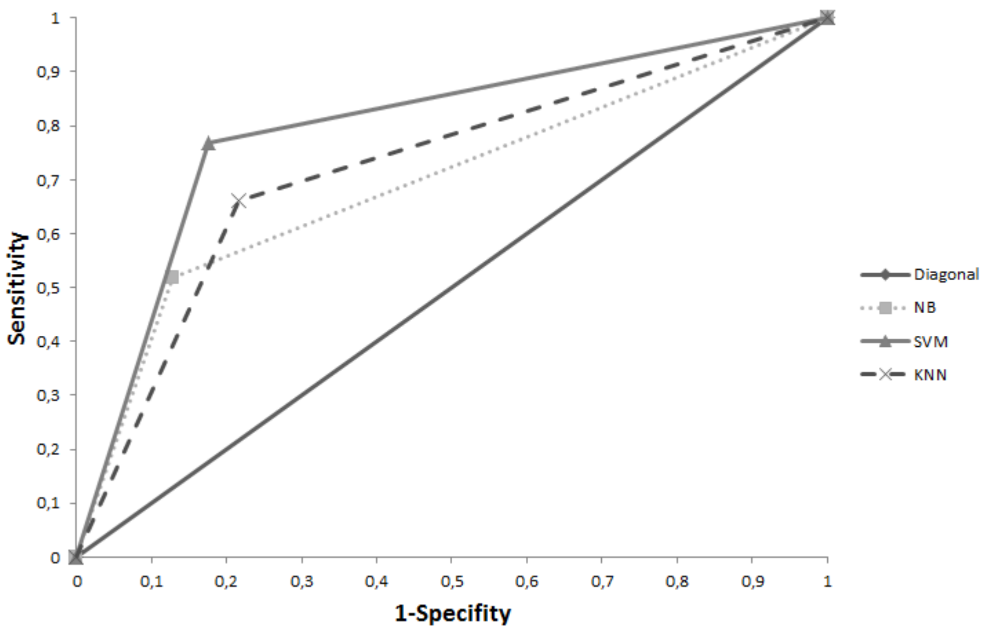

Figura 4. ROC curve with SFE

The graphs and tables expose clearly the superior performance of KPCA applied with any classifier tested, with high accuracy and AUC rates, specifically with KNN and

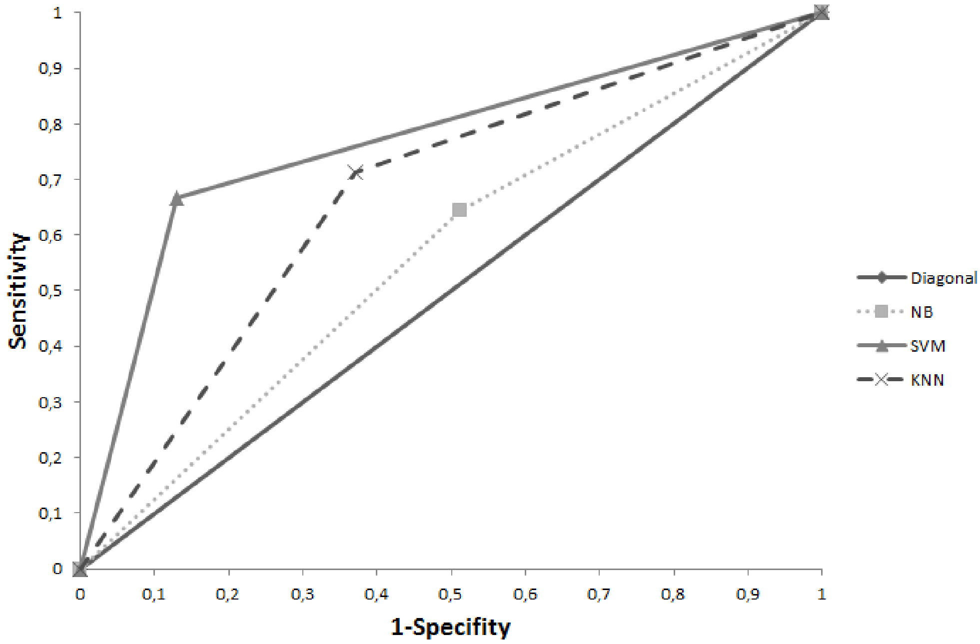

Figura 5. ROC curve with PCA

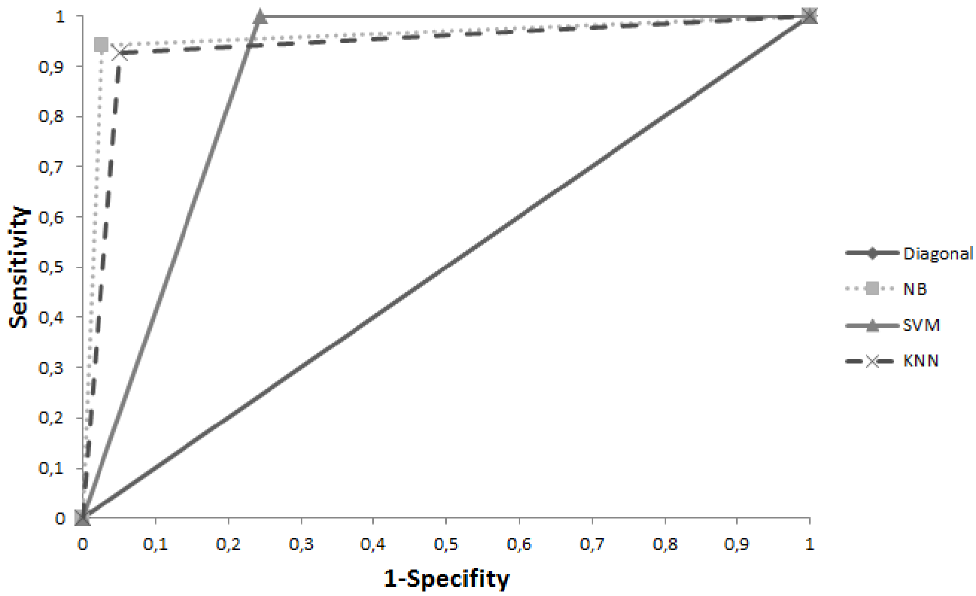

Figura 6. ROC curve with KPCA

Tabela 2. Sensitivity, Specificity and AUC with SFE.

SFE	NB	SVM¨	KNN
Sensitivity	0.519	0.769	0.662
1 - Specificity	0.128	0.175	0.217
AUC	0.648	0.798	0.726

Tabela 3. Sensitivity, Specificity and AUC with PCA.

PCA	NB	SVM¨	KNN
Sensitivity	0.644	0.666	0.712
1 - Specificity	0.512	0.130	0.371
AUC	0.572	0.674	0.671

Tabela 4. Sensitivity, Specificity and AUC with KPCA.

KPCA	NB	SVM¨	KNN
Sensitivity	0.942	1.000	0.927
1 - Specificity	0.026	0.244	0.051
AUC	0.959	0.937	0.958

NB, where produces some good ROC curves, with area higher than 0.95. PCA do not have good results, proving to be insufficient to solve our problem. SFE have some reasonable results both in accuracy and AUC.

So the best combination founded for the problem is a Naïve Bayes or K-Nearest Neighbor classifier using a feature of 13 dimensions produced with Gaussian Kernel PCA, from 17 Haralick texture features in 9 subspaces of Haar Wavelet, which provide a AUC higher than 0.95 and a accuracy of 96% (NB) and 93% (KNN). What is higher than Radiologists [Young and Marrie 1994], how we can see in Table 5.

Tabela 5. Diagnostic Accuracy.

Medical resident	Radiologist	NB with KPCA
66	87	96

4. Conclusion

In this paper, three contemporary machine learning classifiers (Support Vector Machine, K-Nearest Neighbors, and Naïve Bayes) were tested to identify and classify radiographic images in order to to detect and diagnose childhood pneumonia. The classifiers have been evaluated with a dataset taken from clinical routine. The classifiers were optimized, and tested with a cross-validation method to ensure that there is no overfitting. Naïve Bayes and K-Nearest Neighbor have shown best results (96% and 93%, respectively).

In summary, the Naïve Bayes classifier produced most accurate results and has shown to be more stable with this type of images so far. Moreover, it outperforms the best result from previous work, and even outperforms the diagnosis accuracy of Radiologists.

Referências

Abe, S. (2010). *Support Vector Machines for Pattern Classification*. Springer.

Bashar, M. K., Matsumoto, T., and Ohnishi, N. (2003). Wavelet transform-based locally orderless images for texture segmentation. *Pattern Recogn. Lett.*, 24(15):2633–2650.

Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.*, 2(2):121–167.

Cheng, S.-C. (2003). Content-based image retrieval using moment-preserving edge detection. *Image and Vision Computing*, 21(9):809 – 826.

Cherian, T., Mulholland, E., Carlin, J., Ostensen, H., Amin, R., de Campo, M., Greenberg, D., Lagos, R., Lucero, M., Madhi, S., O′Brien, K., Obaro, S., and Steinhoff, M. (2005). Standardized interpretation of paediatric chest radiographs for the diagnosis of pneumonia in epidemiological studies. *Bull. World Health Organ.*, 83(5):353–359.

Cover, T. and Hart, P. (1967). Nearest neighbor pattern classification. *Information Theory, IEEE Transactions on*, 13(1):21–27.

Depeursinge, A., Iavindrasana, J., Hidki, A., Cohen, G., Geissbuhler, A., Platon, A., Poletti, P.-A., and Müller, H. (2012). Comparative performance analysis of state-of-the-art classification algorithms applied to lung tissue categorization. *Journal Digit Imaging*, 23(1):1830.

Doi, K., MacMahon, H., Katsuragawa, S., Nishikawa, R. M., and Jiang, Y. (1999). Computer-aided diagnosis in radiology: potential and pitfalls. *European Journal of Radiology*, 31(2):97 – 109.

Guyon, I. and Elisseeff, A. (2003). An introduction to variable and feature selection. *J. Mach. Learn. Res.*, 3:1157–1182.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: an update. *SIGKDD Explor. Newsl.*, 11(1):10–18.

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

Hotelling, H. (1933). Analysis of complex statistical variables into principal components. *Journal of Educational Psychology*, 24(6):417–441.

Huang, J. and Ling, C. (2005). Using auc and accuracy in evaluating learning algorithms. *Knowledge and Data Engineering, IEEE Transactions on*, 17(3):299–310.

Huang, P. W. and Dai, S. K. (2004). Design of a two-stage content-based image retrieval system using texture similarity. *Inf. Process. Manage.*, 40(1):81–96.

Kokare, M., Biswas, P. K., and Chatterji, B. N. (2005). Texture image retrieval using new rotated complex wavelet filters. *Trans. Sys. Man Cyber. Part B*, 35(6):1168–1178.

Kokare, M., Chatterji, B. N., and Biswas, P. K. (2004). Cosine-modulated wavelet based texture features for content-based image retrieval. *Pattern Recogn. Lett.*, 25(4):391–398.

Ladha, L. and Deepa, T. (2011). feature selection methods and algorithms. *International Journal on Computer Science and Engineering*, 3(5):1787–1797.

Levine, O. S., Lagos, R., Munoz, A., Villaroel, J., Alvarez, A. M., Abrego, P., and Levine, M. M. (1999). Defining the burden of pneumonia in children preventable by vaccination against haemophilus influenzae type b. *Pediatr. Infect. Dis. J.*, 18(12):1060–1064.

Macedo, S. O. d. and Oliveira, L. L. G. d. (2012). Desenvolvimento de um sistema de auxílio ao diagnóstico de pneumonia na infância utilizando visão computacional. In *Workshop de Visão Computacional*.

Oliveira, L. L. G., e Silva, S. A., Ribeiro, L. H. V., de Oliveira, R. M., Coelho, C. J., and Andrade, A. L. S. S. (2008). Computer-aided diagnosis in chest radiography for detection of childhood pneumonia. *I. J. Medical Informatics*, 77(8):555–564.

Schölkopf, B., Smola, A., and Müller, K.-R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.*, 10(5):1299–1319.

van der Maaten, L. (2013). Matlab toolbox for dimensionality reduction. v0.8.1.

van der Maaten, L. J. P., Postma, E. O., and van den Herik, H. J. (2009). Dimensionality Reduction: A Comparative Review. Technical report, Tilburg University.

WHO (2001). Standardization of interpretation of chest radiographs for the diagnosis of pneumonia in children. Technical report, World Health Organization: Department of Vaccines and Biologicals.

WHO (2012). Pneumonia, fact sheet n°331. Technical report, World Health Organization.

Yao, J., Dwyer, A., Summers, R. M., and Mollura, D. J. (2011). Computer-aided diagnosis of pulmonary infections using texture analysis and support vector machine classification. *Academic Radiology*, 18(3):306–14.

Young, M. and Marrie, T. J. (1994). Interobserver variability in the interpretation of chest roentgenograms of patients with possible pneumonia. *Arch Intern Med*, 154:2729–32.