Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Computer Science 18 (2013) 2579 – 2582

Procedia Computer Science

International Conference on Computational Science, ICCS 2013

Comparative performance analysis of machine learning classifiers in detection of childhood pneumonia using chest radiographs

Rafael T. Sousa[a,*], Oge Marques[b], Fabrizzio Alphonsus A. M. N. Soares[a], Iwens I. G. Sene Jr[a], Leandro L. G. de Oliveira[a], Edmundo S. Spoto[a]

[a]*Instituto de Informática Universidade Federal de Goiás (UFG) Goiânia, GO - Caixa Postal 131 74.001-970 Brazil*
[b]*Department of Computer & Electrical Engineering and Computer Science Florida Atlantic University (FAU) - Boca Raton, FL 777 Glades Road 33431 USA*

Abstract

This work extends *PneumoCAD*, a Computer-Aided Diagnosis system for detecting pneumonia in infants using radiographic images [1], 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). Results of our experiments demonstrate that the SVM classifier produces the best overall results.

Keywords: Machine Learning; Computer-Aided Diagnosis (CAD); Childhood pneumonia;

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, most in South Asia and sub-Saharan Africa. This number is higher than the mortality rate for several other diseases, such as AIDS, malaria and tuberculosis, combined [2].

Currently the best and most widely accepted imaging modality for detecting pneumonia is chest radiographs [3]. However, some studies have shown that errors are common in the interpretation of chest radiographs, due to inter-observer variation [4]. 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 [5]. 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. Using feature extraction and supervised learning classifiers a CAD system can mimic the specialist's vision and ability to diagnose.

In this work we use the same features and dataset employed in previous studies [1] [6], 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

*Corresponding author. Tel.: +55-66-84090416.
E-mail address: rafaelsousa@inf.ufg.br.

Selection and peer review under responsibility of the organizers of the 2013 International Conference on Computational Science
doi:10.1016/j.procs.2013.05.444

[6]. *PneumoCAD* is currently in prototype stage. 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 machine learning classifiers combined with feature selection algorithms, to improve *PneumoCAD* accuracy and find out the best classifier for childhood pneumonia detection.

2. Methods

The image dataset used in *PneumoCAD* 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 1 shows examples of the images in the dataset. These images were analyzed by two trained radiologists according to WHO guidelines [7] [8] which produced the ground truth needed to test the machine learning classifiers used in this work.

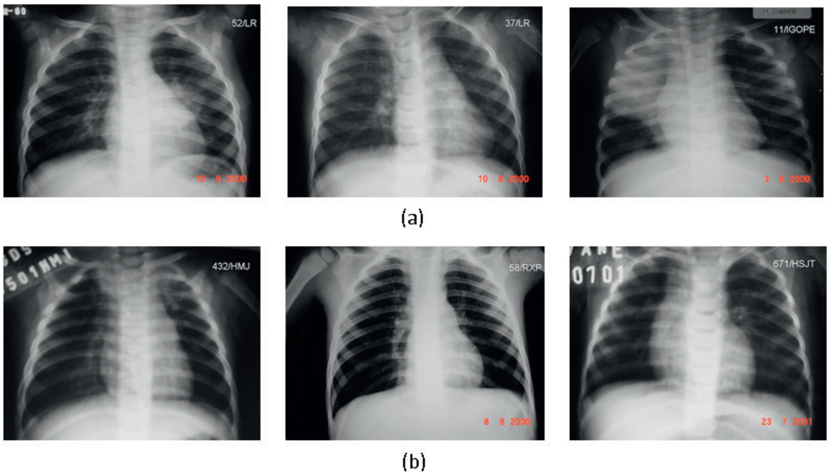

Fig. 1. (a) Three chest radiographs labeled as 'positive' for pneumonia. (b) Three healthy children radiographs.

The following texture-based features were selected, implemented, and tested: 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 [9] [10] [11] [12]. All features have been extracted in nine subspaces of Haar wavelet.

Method 1

First each feature was tested with three classifiers, KNN, SVM and Naïve Bayes. In order to compare with previous results [6], we used three random sets of 40 images each and then, from each set, a random subset of 15 images. We then perform three tests alternating the subsets, two as training set and the other as test. All the three tests are performed only with difference variance feature and with k parameter in KNN defined as 9. For SVM we used standards Gaussian kernel parameters ($C = 1, \sigma = 1$).

Method 2

In order to test the robustness of each classifier applied in pneumonia detection and to avoid overfitting, we applied a second methodology with no randomization of the training set. 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, were removed. We then performed a 10-fold cross-validation test with each classifier using every feature in 9 subspaces of wavelet. After identifying and selecting the three best possible features for each classifier, we performed with these an exhaustive search, testing many possible values for each parameter, according to Table 1.

Table 1. Parameters and intervals.

Classifier	Parameter	Ranges	Step(step size)
KNN	k	[0; 50]	lin(1)
SVM	C, σ	[1; 50],[0;10]	lin(2), lin(0.1)
Naïve Bayes	-	-	-

Method 3

Finally we tested another approach for feature selection, which takes the full feature vector, with all 17 texture features, and uses a Sequential Forward Elimination (SFE) test, which is a simple greedy search algorithm, to find the best feature set for each classifier [13] [14].

3. Results

Method 1

With the first methodology good accuracy results were obtained, and these are the best figures: 82% with KNN, 80% with SVM and 60% with Naïve Bayes. SVM and KNN classifiers obtained good results, but also bad results in other runs, e.g., 50% with KNN and SVM. So a second methodology is important to evaluate the problem using all data set with cross-validation tests, because the random set selection may lead to biases that favor one classifier over another and make the overall comparison less meaningful.

Method 2

Feature tests were performed with the three classifiers. Several features have shown overall good results, namely: entropy, difference variance, sum entropy, and suavity.

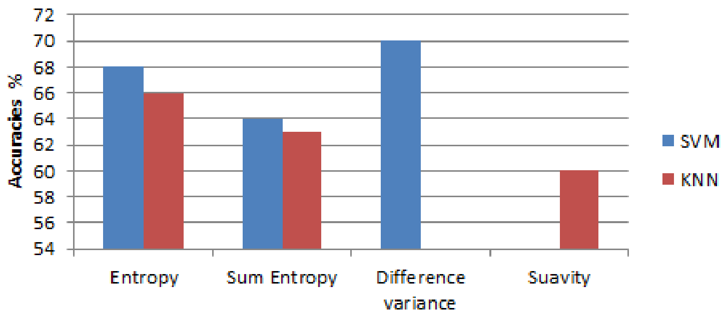

Fig. 2. Optimized SVM and KNN accuracies with each three selected feature.

After obtaining the three best features for KNN and SVM, we performed a calibration test, to improve the accuracy results, as shown on Figure 2. In SVM tests the best values for the parameters were: $\{C = 1.2; \sigma = 1\}$ for entropy, $\{C = 5.7; \sigma = 0.5\}$ for difference variance and $\{C = 0.6; \sigma = 0.6\}$ for sum entropy. For KNN the best calibration led to k=11 for suavity, k=7 for sum entropy and k=9 for entropy.

Method 3

Table 2. Classifiers result accuracies with each features set selected by SFE.

SVM	KNN	PNB
77	70	68

Our third method, using a feature selection algorithm, produced some good results (Table 2). With SVM the selected features were: correlation, average deviation, difference variance and standard deviation, one wavelet subspace from each. With KNN the selected features were: energy and suavity, with one subspace each. Naïve Bayes best result was with: entropy, difference variance and sum average, one subspace each.

So the best combination founded for the problem is a SVM classifier calibrated with $\{C = 1; \sigma = 2\}$ and a feature set selected by SFE algorithm in the Haralick texture features, which provide an accuracy of 77%, what is higher than previous version of PneumoCAD and Medical Residents [4], as shown in Table 3.

Table 3. Diagnosis Accuracy.

Medical Residents	PneumoCAD - KNN without SFE	PneumoCAD - SVM with SFE
66	66	77

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 with best features, and tested with a cross-validation method to ensure that there is no overfitting. SVM and KNN have shown good results (77% and 70%, respectively), but SVM produces a slightly better result in average accuracy. The Naïve Bayes classifier came third, with best accuracy at 68%.

In summary, the SVM classifier produced most accurate results and has shown to be more stable with training data variation. Moreover, it outperforms the best result from previous work, and even outperforms the diagnosis accuracy of medical residents.

References

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

[2] Pneumonia, fact sheet n331, Tech. rep., World Health Organization (2012).

[3] Standardization of interpretation of chest radiographs for the diagnosis of pneumonia in children, Tech. rep., World Health Organization: Department of Vaccines and Biologicals (2001).

[4] M. Young, T. J. Marrie, Interobserver variability in the interpretation of chest roentgenograms of patients with possible pneumonia, Arch Intern Med 154 (1994(23)) 2729–32.

[5] K. Doi, H. MacMahon, S. Katsuragawa, R. M. Nishikawa, Y. Jiang, Computer-aided diagnosis in radiology: potential and pitfalls, European Journal of Radiology 31 (2) (1999) 97 – 109. doi:10.1016/S0720-048X(99)00016-9.
URL `http://www.sciencedirect.com/science/article/pii/S0720048X99000169`

[6] S. O. d. Macedo, L. L. G. d. Oliveira, Desenvolvimento de um sistema de auxílio ao diagnóstico de pneumonia na infância utilizando visão computacional, in: Workshop de Visão Computacional, 2012.

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

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

[9] R. M. Haralick, K. Shanmugam, I. Dinstein, Textural Features for Image Classification, Systems, Man and Cybernetics, IEEE Transactions on SMC-3 (6) (1973) 610–621. doi:10.1109/TSMC.1973.4309314.
URL `http://dx.doi.org/10.1109/TSMC.1973.4309314`

[10] P. W. Huang, S. K. Dai, Design of a two-stage content-based image retrieval system using texture similarity, Inf. Process. Manage. 40 (1) (2004) 81–96. doi:10.1016/S0306-4573(02)00097-3.
URL `http://dx.doi.org/10.1016/S0306-4573(02)00097-3`

[11] M. Kokare, P. K. Biswas, B. N. Chatterji, Texture image retrieval using new rotated complex wavelet filters, Trans. Sys. Man Cyber. Part B 35 (6) (2005) 1168–1178. doi:10.1109/TSMCB.2005.850176.
URL `http://dx.doi.org/10.1109/TSMCB.2005.850176`

[12] M. K. Bashar, T. Matsumoto, N. Ohnishi, Wavelet transform-based locally orderless images for texture segmentation, Pattern Recogn. Lett. 24 (15) (2003) 2633–2650. doi:10.1016/S0167-8655(03)00107-7.
URL `http://dx.doi.org/10.1016/S0167-8655(03)00107-7`

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

[14] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, J. Mach. Learn. Res. 3 (2003) 1157–1182.
URL `http://dl.acm.org/citation.cfm?id=944919.944968`