Measuring the relative extent of pulmonary infiltrates by hierarchical classification of patient-specific image features

S Tsevas[1,2] **and D K Iakovidis**[1,3]

[1] Department of Computer Science, University of Geneva, Geneva, Switzerland
[2] Department of Informatics and Computer Technology, Technological Educational Institute of Lamia, Lamia, Greece

E-mail: s.tsevas@ieee.org and dimitris.iakovidis@ieee.org

Abstract

Pulmonary infiltrates are common radiological findings indicating the filling of airspaces with fluid, inflammatory exudates, or cells. They are most common in cases of pneumonia, acute respiratory syndrome, atelectasis, pulmonary oedema and haemorrhage, whereas their extent is usually correlated with the extent or the severity of the underlying disease. In this paper we propose a novel pattern recognition framework for the measurement of the extent of pulmonary infiltrates in routine chest radiographs. The proposed framework follows a hierarchical approach to the assessment of image content. It includes the following: (a) sampling of the lung fields; (b) extraction of patient-specific grey-level histogram signatures from each sample; (c) classification of the extracted signatures into classes representing normal lung parenchyma and pulmonary infiltrates; (d) the samples for which the probability of belonging to one of the two classes does not reach an acceptable level are rejected and classified according to their textural content; (e) merging of the classification results of the two classification stages. The proposed framework has been evaluated on real radiographic images with pulmonary infiltrates caused by bacterial infections. The results show that accurate measurements of the infiltration areas can be obtained with respect to each lung field area. The average measurement error rate on the considered dataset reached 9.7% ± 1.0%.

Keywords: image processing, pattern recognition, medical imaging, chest radiographs, feature extraction, classification, pulmonary infiltrates

1. Introduction

Medical imaging comprises an exceptional non-invasive approach to screening of internal structures and processes of the human body, while providing measurable evidence regarding volumes [1], densities [2], motion and flows [3]. Such evidence could provide substantial cues in the differential diagnosis and monitoring of the progress of diseases.

In this light, radiographic imaging of the chest can be useful to the screening and measurement of fluid, inflammatory exudates or cells, filling the airspaces of the lungs. These are usually caused by diseases such as pneumonia, acute respiratory syndrome, atelectasis, pulmonary oedema and haemorrhage. A common radiographic manifestation of these findings is known as pulmonary infiltrates, which appear as whiter opacities within the lung fields. The grey-level intensity of these opacities is higher than the opacities of the lung parenchyma, which is normally filled with air, and lower than the intensity of the bones, which are generally denser anatomic structures. The extent of these findings is usually correlated

---

[3] Author to whom any correspondence should be addressed.

with the extent or the severity of the disease. For example, in the case of pulmonary infections the infiltrates, known as consolidations, spread as the disease progresses over time [4, 5].

The presence and the extent of a patient's pulmonary infiltrates are assessed by visual examination of his/her chest radiograph. The assessment is primarily made based on the grey-level intensity of the infiltrates in relation to the intensities of the surrounding structures. Intensity is an image feature correlated with the density of the depicted objects [4]. The interpretation of a chest radiograph requires both knowledge and experience from the person evaluating its content; however it is affected by his/her subjectivity. Therefore it is widely accepted that a method for automatic analysis of chest radiographs would be a great asset to the medical community, contributing to the objectification of their interpretation. Computational methods for automatic analysis of chest radiographs have already been proposed in several studies. Such methods include detection of the lung fields [6], detection of the ribs [7], lung nodule detection [8], whereas other methods cope with the detection or assessment of abnormalities associated with the presence of pulmonary infections [9, 10]. In [9] a k-nearest-neighbour (k-NN) approach has been considered for the classification of lung field patterns into normal and abnormal ones caused by tuberculosis, which is a pulmonary infection of bacterial origin. In that study the patterns were represented by Gabor features estimated as first-order statistics of Gabor-filtered images. In [10] three classification methods, namely C4.5, neural networks and CART, have been evaluated for the detection of patterns of pneumonia and severe acute respiratory syndrome using feature vectors of both first- and second-order statistical measures. Recently, in another study wavelet features have been utilized instead of Gabor-based features to obtain a less computationally complex feature extraction phase for the detection of radiographic patterns of childhood pneumonia under a nearest-neighbour classification framework [11].

However, most of these studies have not taken into consideration the diversity of the chest radiographs regarding the different acquisition settings or the different radiographic characteristics of the examined bodies. Some of them claim the use of identical radiographic imaging devices and settings for the acquisition of the considered datasets. Although this could be a way to cope with the diversity issue, the intensities of the same anatomic structures depicted in a chest radiograph can vary between different patients. For example, the mean intensity of a normal lung field in a chest radiograph of an overweight patient is usually higher than the mean intensity of a thinner patient's lung field. This is due to a thinner layer of fat, the intensity of which is superimposed on the intensities of the underlying lung field. Therefore, an image normalization method that would reduce this effect, leading to comparable grey-level intensities for the same specimens across different chest radiographs, would positively impact image analysis methods, especially the supervised ones that utilize features extracted from multiple chest radiographs for machine learning purposes.

Many studies also consider image texture as an important feature for the discrimination of chest infiltrates [9, 11]. Motivated by these studies, in [12] we proposed an unsupervised methodology for the discovery of infiltrates associated with bacterial pulmonary infections. The methodology was based on a hierarchical scheme of partitional clustering fusing information represented by both intensity and textural image features. This clustering approach was applied on each chest radiograph separately, thus avoiding the need for a training set of radiographs which would require proper normalization [4].

However, the use of unsupervised methodologies requires a predefined number of clusters that hinders their use for the detection of pulmonary infiltrates. For example, aiming to detect such infiltrates in a chest radiograph with an unsupervised clustering algorithm, one would set the target number of clusters to two, so that one cluster would contain the infiltrates and one cluster would contain the normal samples. This algorithm has no prior knowledge of what is defined as normal or infiltrate, therefore it would assign samples to both clusters based only on their pairwise similarities. The samples that would be assigned (inevitably) to the infiltrate cluster would be false positive samples. Thus the requirement for supervised methodologies not only capable of assessing the extent of an existing infection but also able to detect it by including prior knowledge is prevalent. This requirement comes along with the need for proper feature normalization between images previously discussed. The standard normalization techniques such as the z-score and min-max normalization do not take into account any information about the settings used for the acquisition of a chest radiograph or any patient-specific information. Therefore they are expected to be only approximative when it comes to the analysis of diverse sets of chest radiographs acquired with different settings and from patients with different anatomic characteristics [14].

In order to cope with this problem, we propose a novel image normalization technique that takes into account patient-specific characteristics visible in the chest radiographs. More specifically we consider the region of the mediastinum and the spinal cord as a reference region for normalization. The considerations motivating our approach include the following: (a) this region can be easily detected since it lies roughly somewhere in the middle of the radiograph and it is generally characterized by higher intensities than other regions in a chest radiograph [13, 14]; (b) although the absolute intensity features may considerably vary between different chest radiographs, the relative intensities of the different anatomic structures with respect to the spinal cord/mediastinum region are expected to be rather invariant. This is explained by the nature of chest radiographs itself. The radiographic image intensities represent the densities of the depicted anatomic structures, whose relative values are similar for all human patients. Such a patient-specific normalization technique is a tool to cope with the analysis of multiple chest radiographs in a supervised classification environment. Towards this direction, in this paper we extend the unsupervised hierarchical clustering methodology presented in [12] by incorporating

prior knowledge to detect whether pulmonary infiltrates exist or not. Therefore, the contribution of this work, in the context of a clinically realizable system for the measurement of the extent of pulmonary infiltrates, is twofold:

- a patient-specific image normalization technique enabling the application of supervised methodologies for the analysis of diverse sets of chest radiographs, and
- a supervised hierarchical classification methodology capable of detecting pulmonary infiltrates and measuring their extent.

Since the radiographic opacities are the first cues considered in the reading of a chest radiograph by the experts, classification of the normalized intensity feature space is first performed. By computing the distribution over the labels assigned by a k-NN classifier in the k-neighbourhood, this first step results in assigning the new data a probability of belonging to the class representing normal lung parenchyma or to the class representing the infiltrates. The image signatures (feature vectors) classified with a higher uncertainty by the classifier are characterized as ambiguous and their ambiguity is further resolved in the second step of the proposed methodology which involves classification of the texture space. An early version of this approach has been preliminarily studied in [15]. The promising results obtained from that study motivated its extension for the measurement of the relative extent of pulmonary infiltrates and its thorough evaluation in this paper. The proposed methodology and the extended experimental results from its application on real chest radiographs are presented in the following sections.

2. Methodology

The proposed pattern recognition framework is based on a supervised hierarchical classification scheme, extending the unsupervised approach proposed in [12]. Classification is enhanced by a novel image normalization technique which results in patient-specific features used for the discrimination of pulmonary infiltrates from normal parenchyma in plain chest radiographs. The radiographs are first normalized using the mediastinum/spinal cord region as a reference. In the following, the lung fields are isolated in regions of interest (ROIs) defined either manually or with a pre-processing lung field boundary detection algorithm [13, 14]. The proposed image analysis methodology is applied only in these ROIs. Finally, the extent of the infiltrates is measured as a ratio of the number of pixels that belong to infiltrate areas over the total number of pixels that belong to the area of the lung parenchyma. The proposed radiograph normalization scheme along with the extraction of the patient-specific features is described in sub-section 2.1 and the details of the classification methodology are presented in sub-section 2.2.

2.1. Patient-specific feature extraction

2.1.1. Extraction of grey-level histogram signatures. In order to extract patient-specific intensity signatures the radiograph is pre-processed by histogram equalization and normalized using
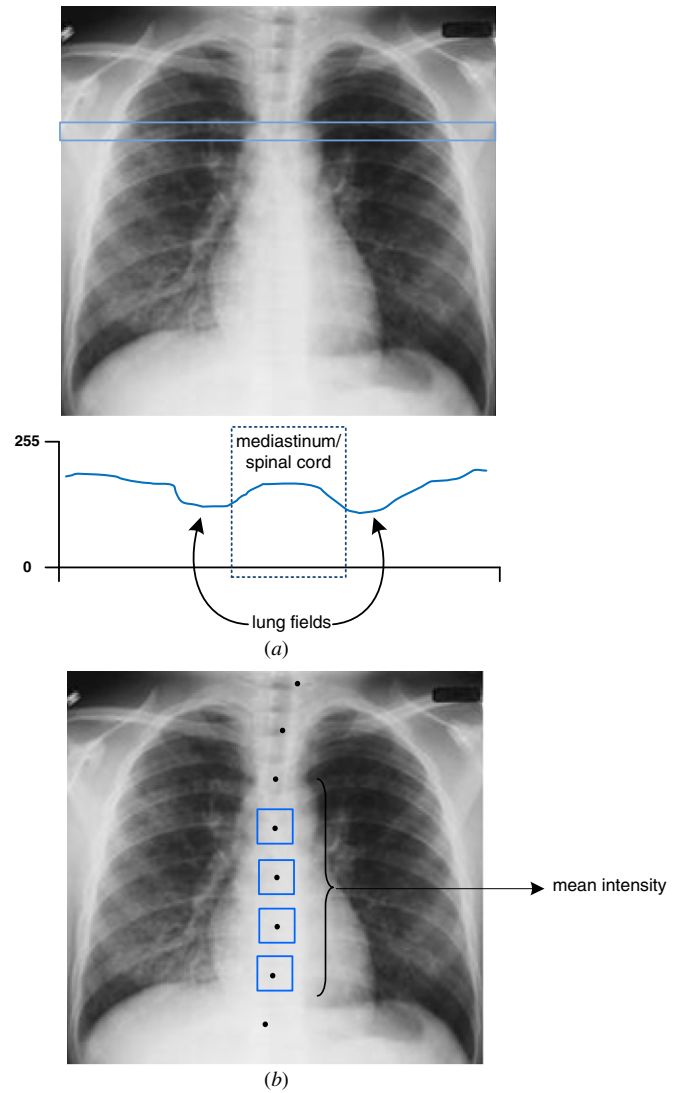

Figure 1. (*a*) An example of a horizontal image profile; (*b*) the detected mediastinum/spinal cord maxima are indicated as black points. Samples are acquired at the detected points and their mean intensity is then calculated.

as a reference the mediastinum/spinal cord region. Histogram equalization of each radiographic image is performed so that each image would contain a uniform distribution of intensities. Considering a grey-scale image I with m_k being the number of occurrences of the grey-level k, the probability of occurrence of a pixel with grey-level k in the image would be:

$$p_I(k) = \frac{m_k}{m} \qquad (k = 0, 1, 2, \ldots, L-1), \qquad (1)$$

where m is the total number of pixels in the image and L is the total number of grey-levels in the image. Histogram equalization is actually a transformation of image I to a new image such that the cumulative distribution function that corresponds to $p_I(k)$ would be linearized across its value range.

After histogram normalization and in order to detect the region of the mediastinum and the spinal cord, each radiograph is uniformly sampled from top to bottom using non-overlapping rectangular sub-images that were further

processed according to [13, 14]. For each sample an average horizontal profile, i.e. the mean grey-level of its rows, per column is estimated. Horizontal profiles of consecutive image samples are utilized in order to obtain spatial instances of the radiograph from which the mediastinum area is detectable even if the patient is bent (figure 1(*a*)).

Each profile is smoothed by following a moving average approach so that noise insensitivity is obtained [13]. The localization of the maxima of each profile remains practically unaffected by the presence of infiltrates, since their density, and therefore their intensity, is generally lower than the density of mediastinum. Mediastinum and spinal cord points are found by classification of the profile maxima into three sets: (a) spinal cord points, (b) points of the left side of the ribcage and (c) points of the right side of the ribcage [14]. The mediastinum region is then sampled at the detected points. The selection of the sample size was based on the image size so as not to stray from the mediastinum's region (figure 1(*b*)).

Image normalization is performed using the central tendency (i.e. mean or median) of the intensity signatures at the detected mediastinum/spinal cord samples as a normalizing factor of the pixel intensities in the radiograph. The steps towards normalization are summarized in figure 2.

The outcome of the proposed normalization approach is to make image intensities comparable across the whole dataset. Thus the intensities that correspond to the normal lung areas should have similar ranges over the different radiographs, and the intensities of the infiltrate areas should also have similar ranges over the different radiographs that form the dataset. In addition, the range of the normal lung intensities should be as far from the range of the infiltrate intensities as possible. An example of the effect of the patient-specific normalization on two chest radiographs from two different datasets is illustrated in figure 3. The radiograph illustrated in figure 3(*a*) comes from the dataset described in section 3 of this study, whereas the image illustrated in figure 3(*b*) comes from the publicly available JSRT dataset [16]. It should be noted that the two images differ in quality and have been acquired using different devices. However, after normalization the intensity signatures of the respective regions of the lung fields fall within the same range.

Figure 4 illustrates the effect of patient-specific normalization on the intensities of two radiographs that demonstrate infiltrate areas. As can be observed, the histograms of the selected infiltrate areas fall within the same range. However, it should be noted that although normal lung parenchyma areas yield comparable intensity signatures after their patient-specific normalization, the same does not always hold for areas that demonstrate infiltrates. This happens because infiltrates of different density, size and degree may occur, depending on the aetiology and the extent of the infection.

2.1.2. Extraction of textural signatures. A Gabor filter is a band-pass spatial filter with selectivity to both orientation and spatial frequency. Gabor filter features are suitable for detecting local structural patterns from images, whereas they have been reported as excellent texture descriptors in various
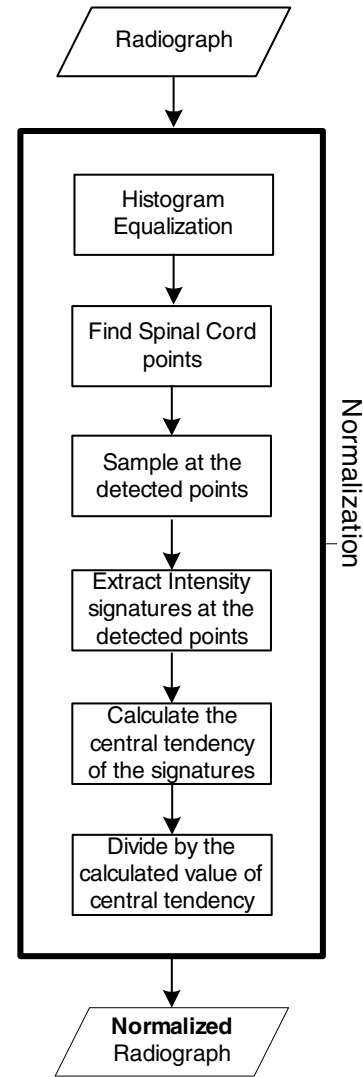

Figure 2. Patient-specific chest radiograph normalization using mediastinum/spinal cord sampling.

studies [17–23]. In order to obtain a Gabor-filtered image, each image $I(x, y)$ is convolved with a two-dimensional Gabor filter function $g(x, y)$:

$$r(x, y) = \iint\limits_{\Omega} I(\xi, \eta) g(x - \xi, y - \eta) \, \mathrm{d}\xi \, \mathrm{d}\eta. \qquad (2)$$

We used the following family of Gabor functions [18, 23]:

$$g_{\lambda, \Theta, \varphi}(x, y) = \mathrm{e}^{-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}} \cos\left(2\pi \frac{x'}{\lambda} + \varphi\right), \qquad (3)$$

where

$$x' = x \cos(\Theta) + y \sin(\Theta) \quad \text{and}$$
$$y' = -x \sin(\Theta) + y \cos(\Theta). \qquad (4)$$

In equation (3) σ stands for the standard deviation of the Gaussian factor, γ represents the spatial aspect ratio, λ^{-1} stands for the spatial frequency of the harmonic factor, Θ is the oscillation orientation and finally, the parameter φ determines the symmetry of the Gabor function. For $\varphi = 0$ and $\varphi = \pi$ the function is symmetric with respect to the centre point $(0, 0)$; for $\varphi = \pm 0.5\pi$ the Gabor function is anti-symmetric.
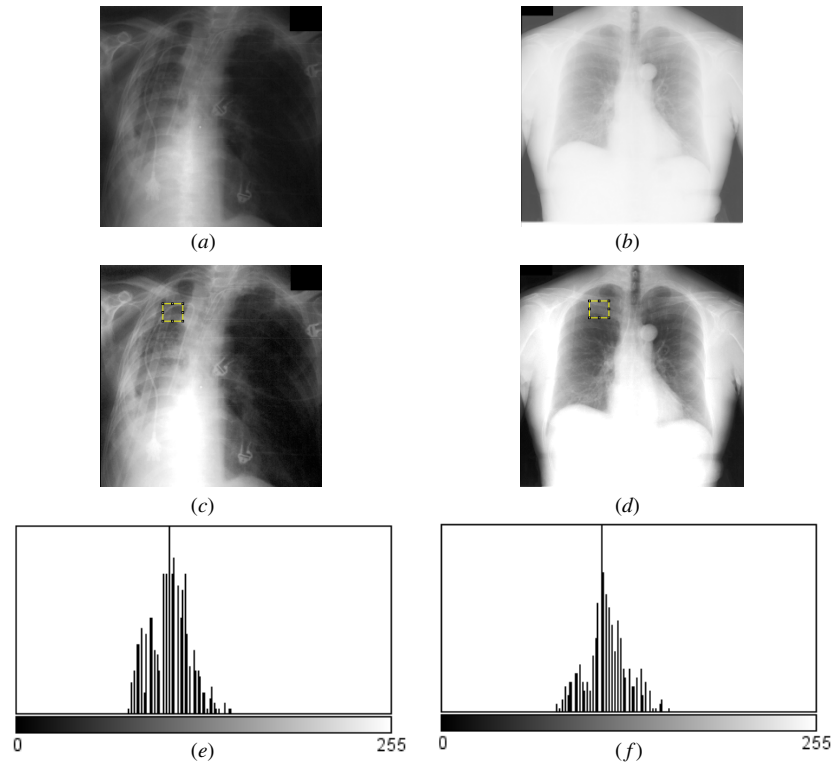
Figure 3. The effect of patient-specific normalization on the intensities of the normal lung parenchyma in two chest radiographs from different datasets. (*a*) Original chest radiograph from the dataset considered in our experiments. (*b*) Original chest radiograph from the JSRT dataset [16]. (*c*) Patient-specific normalization of (*a*). (*d*) Patient-specific normalization of (*b*). (*e*) Grey-level histogram of the ROI in (*c*) with mean 104 ± 13. (*f*) Grey-level histogram of the ROI in (*d*) with mean 112 ± 14.
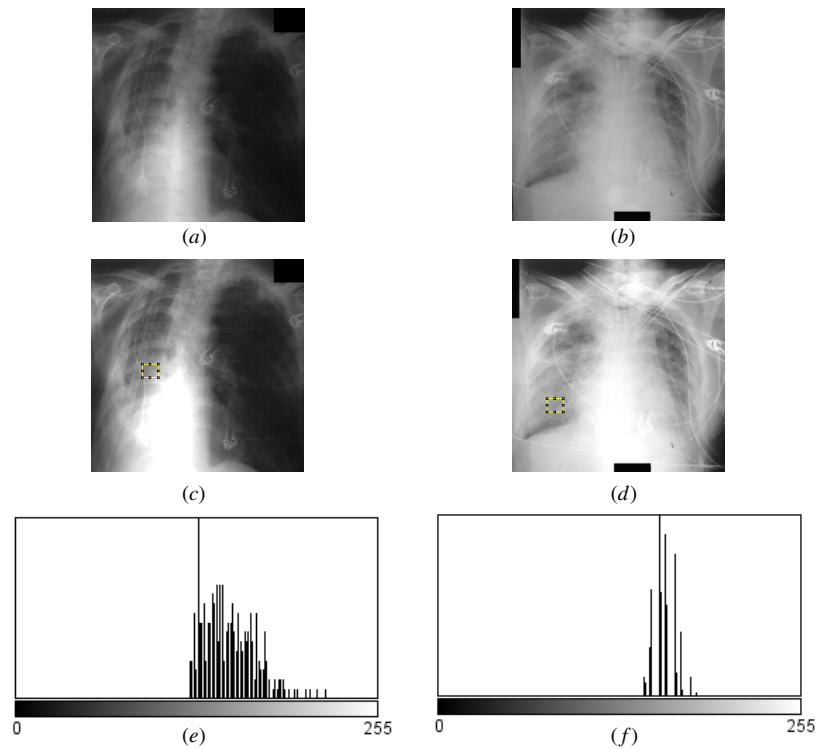

Figure 4. The effect of patient-specific normalization on the intensities of the infiltrates in two chest radiographs from the dataset considered in our experiments. (*a*) The chest radiograph from figures 3(*a*). (*b*) Another chest radiograph from the available dataset. (*c*) Patient-specific normalization of (*a*), (*d*). Patient-specific normalization of (*b*). (*e*) Grey-level histogram of the ROI in (*c*) with mean 151 $\pm$ 18. (*f*) Grey-level histogram of the ROI in (*d*) with mean 159 ± 7.

In this study Gabor energies have been considered so as to represent radiographic image texture. This choice has been motivated by their robustness to noise. Noise may be introduced mainly during the image acquisition process. Gabor energy feature is related to a model of complex cells in the primary visual cortex [24] and is derived by combining in each image point the output of the symmetric and anti-symmetric kernel filters into a single quantity:

$$GE_{\lambda,\Theta}(x, y) = \sqrt{r_{\lambda,\Theta,\phi=0}^2(x, y) + r_{\lambda,\Theta,\phi=-0.5\pi}^2(x, y)}. \quad (5)$$

The combination of symmetric and anti-symmetric Gabor filters in equation (5) results in a new nonlinear filter bank with the same coverage of the spatial frequency domain.

The patient-specific normalization algorithm proposed in this study could be applied to other medical imaging domains as well so as to lead to patient-specific image features. This is feasible once other salient points are considered instead of the mediastinum/spinal cord points. For example, in thyroid US imaging, such points could be sampled across the hyper-echoic lines bounding the thyroid lobe [25].

2.2. Hierarchical classification

The proposed classification methodology is supervised and therefore it has two phases: a training and a testing phase. The training phase requires that the physicians indicate the infiltration areas within the lung fields by graphically annotating multiple chest radiographs of patients with diagnosed pneumonia. Local grey-level histogram and Gabor energy signature sets are extracted from non-overlapping square sub-images raster-sampled from the lung area, as described in the previous sub-sections. These sets comprise the training data representing the prior knowledge required for supervised classification. This is implemented by the k-NN algorithm mainly because it is less parametric than other classifiers while being nonlinear [26].

The testing phase involves the analysis of new chest radiographs, not previously used for classifier's training. It aims to classify samples extracted from the lung fields into two classes, representing pulmonary infiltrates and normal lung parenchyma. The area covered by the samples corresponding to the pulmonary infiltrates is considered for the measurement of the extent of infiltrates.

The testing phase proceeds in two steps. Initially, normalized local grey-level histogram signatures (section 2.1.1) capturing image intensity information are extracted from non-overlapping square sub-images raster-sampled from the lung area. The empirical distribution over class labels is computed by normalizing the counts for each class leading to a probabilistic interpretation of the classification result. Such a probabilistic classifier returns a probability distribution over classes for each signature feature vector according to the equation

$$P(y|x) = \frac{1}{k} \sum_{x_i \in N_k(x)} y_i \qquad P(y|x) \in [0, 1], \quad (6)$$

where x is a point of the test space, $N_k(x)$ is the neighbourhood of x defined by the k closest points x_i in the training sample and

$P(y|x)$ is the resulting probability of a test point x to belong to the class y_i. A probability threshold (T) is used only in the first step of the proposed methodology as a certainty level that determines that a sub-image can be classified as normal or as demonstrating infiltration. If the resulting probability $P(y|x)$ of a test signature to belong to a specific class is higher than a threshold T, the certainty of this signature to belong to this class is high. The system classifies the signature as representing normal lung parenchyma or infiltrate areas, otherwise the corresponding sub-image is considered as ambiguous. Thus, the ambiguous class represents the sub-images that demonstrate a certainty level that is lower than the threshold. Their ambiguity is further resolved in the second step of the proposed methodology which involves classification of the texture space.

The need for a class representing ambiguous samples that have to be further declared stems from the fact that, in contrast to normal parenchyma areas, infiltrate areas do not always yield comparable intensity signatures after their patient-specific normalization. As was mentioned in the previous sub-section, this is due to the different density, size and degree that an infiltrate may demonstrate, depending on the aetiology and the extent of the infection. Thus, the ability to discriminate the class of a sample in this case is now more dependent on its textural properties than on their mean grey-level or variance. For this reason, the proposed framework takes into consideration the textural features of the radiograph as well, so as to enhance its ability to discriminate normal from infiltrate patterns for those patterns whose classification is not possible in only the first step.

Therefore, as a second step, textural features are extracted from the ambiguous sub-images yielding a texture signature for each sub-image. These signatures are formed by the energies estimated from the outputs of two Gabor filter banks: one with symmetric and one with anti-symmetric Gabor kernels as was described in section 2.1 [18, 19]. The textural signatures are subsequently classified using the k-NN algorithm in its simple binary form.

The classification results of the two steps are aggregated. The sub-images classified as normal lung parenchyma and the sub-images classified as infiltrates from both steps are merged together to form the final classification result. A block diagram describing the proposed methodology is illustrated in figure 5.

The existence of sub-images classified as representing infiltrates denotes the existence of a possible infection. The relative extent of the infection can be calculated as the ratio R of the pixels corresponding to infiltrates (n_{inf}) to the total number of pixels of the corresponding lung fields (n):

$$R = \frac{n_{\text{inf}}}{n} \cdot 100\%. \quad (7)$$

3. Experimental evaluation

3.1. Description of the dataset

Extensive experiments were conducted for the evaluation of the proposed methodology. The dataset used in this study is composed of chest radiographs obtained from 144 patients
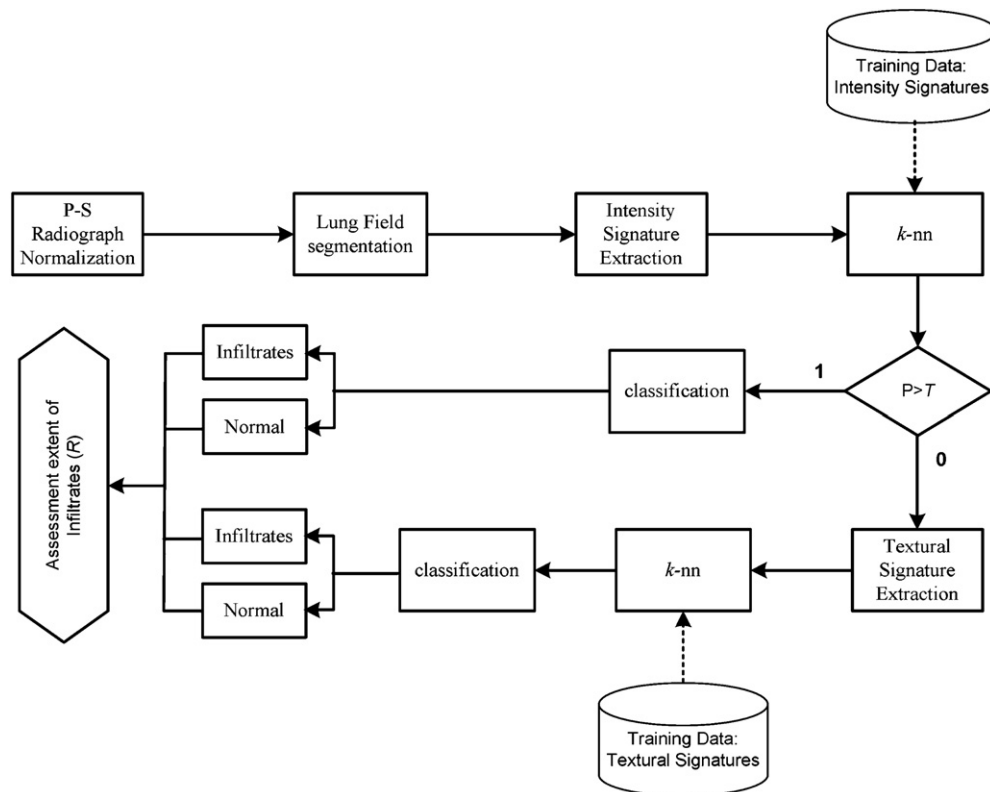
Figure 5. Block diagram describing the proposed methodology. P-S stands for patient specific.

Table 1. Radiograph normalization methods.

Method no	Description
1	Raw images (without any normalization)
2	Histogram equalization
3	Normalization by dividing with mediastinum's mean intensity
4	Histogram equalization followed by normalization by dividing with mediastinum's mean intensity
5	Normalization by subtracting mediastinum's mean intensity
6	Histogram equalization followed by normalization by subtracting mediastinum's mean intensity
7	Contrast stretching by min-max normalization

hospitalized in intensive care unit with pulmonary infections manifested as infiltrates. The chest radiographs have been acquired from different x-ray devices using different image acquisition settings. They have been digitized at a spatial resolution of 2k × 2k pixels with grey-level depth of 8 bits per pixel, and stored in DICOM format. The lung fields were isolated by being delineated by experts, whereas pulmonary infiltrates were localized and delineated. Lung fields were further sampled to 32 × 32 pixel rectangular sub-images leading to a dataset of 29 657 sub-images corresponding to infiltrate areas and 63 069 sub-images for the normal areas. Both intensity and texture features were then calculated for each sub-image.

In order to evaluate the effectiveness of the proposed methodology, extensive experiments were conducted so as to tune the parameters of the system and compare to the state of the art. These experiments have been conducted in two phases.

- Qualitative comparison of the proposed radiograph normalization scheme with alternative ones for the extraction of patient-specific features.

- Assessment of the classification performance of the proposed methodology for different posterior probability thresholds following a leave-one-radiograph-out approach and comparison with the state of the art.

3.2. Evaluation of the normalization scheme

As was described in section 2.1, the need for a specific normalization scheme arose so as to obtain similar intensity ranges for either normal or infiltrate patterns across different radiographs. A successful normalization technique would yield well-separated classes in terms of their intensity signatures. For this reason two requirements were identified to assess the quality of the normalized features: (a) the intensity signatures of the two pattern classes (normal parenchyma and infiltrates) should be as separated as possible, and (b) the intensity signatures of each class should be as similar as possible.

Several image normalization techniques were considered that take into account the patient-specific information provided
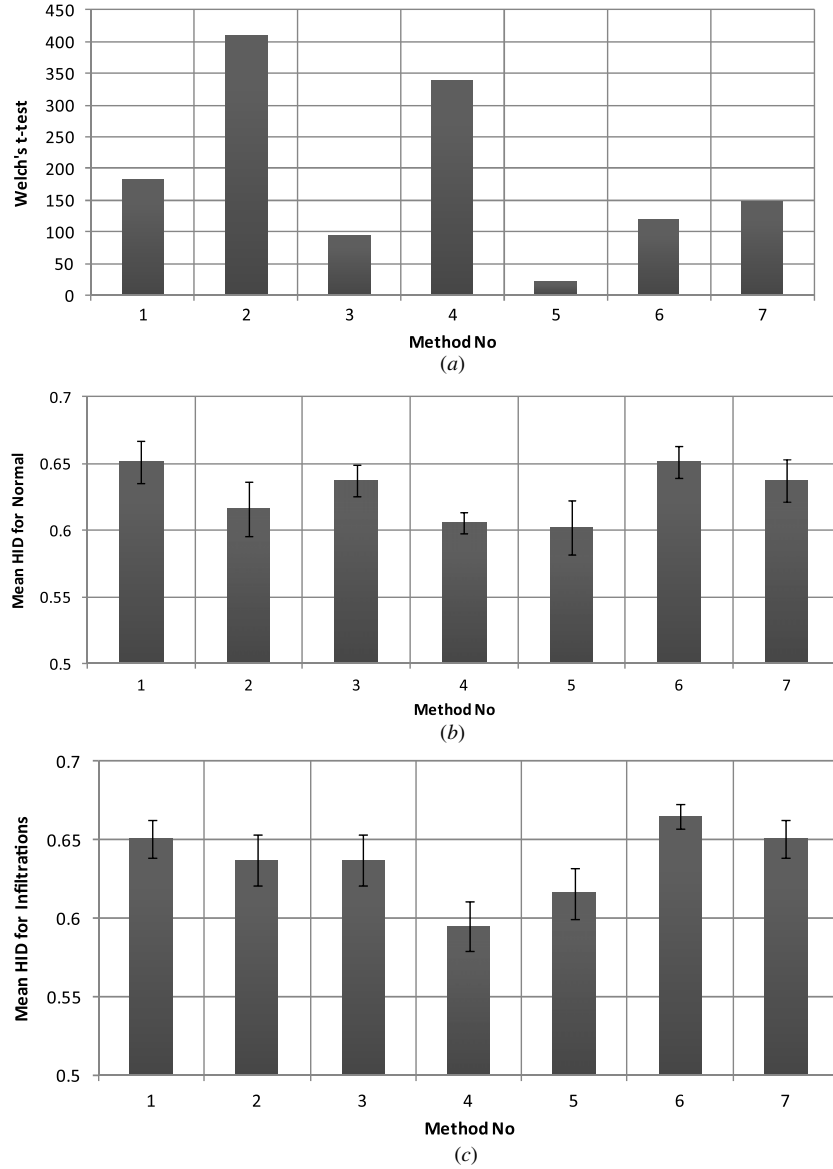
Figure 6. For each one of the methods shown in table 1: (*a*) *t*-value between the samples comprising infiltrate and normal groups, (*b*) mean HID between infiltrate samples, (*c*) mean HID between normal samples. Calculations were made using a set of samples where all infiltrate samples from the dataset were included so as to form the infiltrate group of samples and an equal number of normal samples were randomly selected so as to form the normal group. The error represents two standard deviation units.

by the mediastinum/spinal cord area. The tested techniques are summarized in table 1. For the evaluation of the best normalization scheme in terms of the above-mentioned requirements, Welch's *t*-test was used as a measure of the separation of two sets. In order to measure how similar are the intensity signatures of each class, the mean overlap of the samples' intensity histograms was taken into consideration as it was expressed by the mean histogram intersection distance (HID) [27] between the samples of each class.

Welch's *t*-test [28] is used for testing the null hypothesis that samples in two groups are independent random samples that possibly demonstrate unequal variance:

$$t = \frac{\bar{X}_i - \bar{X}_n}{\sqrt{\frac{s_i^2}{N_i} + \frac{s_n^2}{N_n}}}, \tag{8}$$

where $\bar{X}_i$ and $\bar{X}_n$ demonstrate the group means, s_i and s_n the respective standard deviation and N_i and N_n the group size. Indices i and n are indicative of the class representing the infiltrates and the class representing the normal samples, respectively.

In order to determine the normalization scheme that best meets the requirements discussed earlier, the *t*-test value between the samples that represent the two classes was calculated (figure 6(*a*)). The mean HID was calculated between the samples of each class (figures 6(*b*), (*c*)). The obtained results for each normalization method are illustrated in figure 6.

Figure 6 shows that the normalization technique that best meets the two requirements is the one described in method 4 (histogram equalization of the radiograph followed by normalization by dividing with mediastinum mean). In this
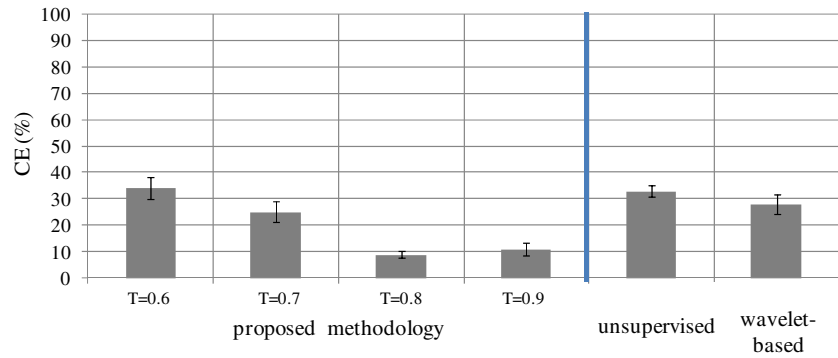
Figure 7. Classification performance of the proposed methodology in terms of average CE for different probability thresholds. The performance of the proposed methodology is compared with the unsupervised [12] and the wavelet-based methodologies [11]. The error-bar represents two standard deviation units.

case the second highest value of t is demonstrated ($t = 337.34$), which indicates good separation between the classes, while at the same time exhibits the smallest mean distance between the members of each class (0.60 for infiltrate group and 0.59 for normal group). Small mean distance means greater overlap between the histograms of the samples of each group.

3.3. Evaluation of the hierarchical classification scheme

The performance of the proposed scheme was evaluated in terms of its pattern classification ability, whereas its measurement capabilities were further compared with the ones of the unsupervised methodology proposed in [12], and the wavelet-based approach presented in [11]. In order to train the classifiers, a training set for the classification was constructed for infiltrate as well as normal regions using a balanced proportion of signatures from normal and infiltrate samples in a way that all abnormal samples were included and an equal number of normal samples was randomly selected. The resulting training set comprised 59 314 labelled features. Moreover, the mean was selected as a more computationally efficient measure of data central tendency since preliminary experimentation demonstrated that both mean and median yield comparable results. The importance of learning from a balanced class distribution was stressed in [29]. In this case the classifiers generally come up with fewer but more accurate classification rules for the minority class than for the majority class. So, such an approach is expected to enhance the classification of abnormal samples and thus increase the system's sensitivity.

In order to proceed with the experiments, the optimal value of the k nearest neighbours was chosen by ten-fold cross-validation employed on the training data. The chosen k-values were those that exhibited the smallest prediction error for each feature set. Thus, for intensity features, the selected value for k was 15 with negligible differences in the system's performance for the range between 13 and 18, whereas 5 nearest neighbours were selected for the textural feature set. Table 2 summarizes the selected k-values.

In order to determine the probability threshold (T) above which a sample is considered as classified in the first step and not ambiguous the performance of the proposed methodology

Table 2. k-values selected.

Feature set	k-value	Prediction error (%)[a]
Intensity	15	4.2 ± 0.31
Texture	5	3.2 ± 0.22

[a] As obtained by ten-fold cross-validation.

for different probability thresholds was assessed in terms of its classification error (CE) [30]:

$$\text{CE} = \left(1 - \frac{\text{tp} + \text{tn}}{\text{tp} + \text{fp} + \text{tn} + \text{fn}}\right) \cdot 100\%, \qquad (9)$$

where tp stands for true positive (the total number of pixels classified as infiltrates by both the methodology under evaluation and the expert), tn for true negative (total number of pixels classified as normal lung parenchyma by both the methodology under evaluation and the expert), fp for false positive (total number of pixels classified as normal lung parenchyma by the expert and as infiltrate by the methodology under evaluation) and fn for false negative (total number of pixels classified as infiltrate by the expert and as normal by the methodology under evaluation).

For this reason, a leave-one-radiograph-out cross-validation scheme was employed on the available dataset. In each iteration of the validation scheme the test set consisted of samples extracted from one radiograph. Furthermore, a set of samples with balanced class distribution was extracted from the rest of the radiographs so as to make the classifier's training set. This was repeated until all radiographs were used for testing. The results obtained by the proposed methodology for the different probability thresholds along with those obtained by the unsupervised scheme proposed in [12] and the wavelet-based methodology proposed in [11] are summarized in figure 7.

As can be observed in figure 7, for a probability threshold equal to 0.8 the lowest CE rate achieved by the proposed methodology is $9.2\% \pm 1.4\%$. For higher thresholds ($T = 0.9$), the CE obtained by the proposed methodology is comparable to the one obtained for $T = 0.8$, however, with a decreasing mean and an increasing variance ($11.1\% \pm 2.4\%$). This is justified by the fact that for such high probability thresholds there is a rise in the number of samples that are
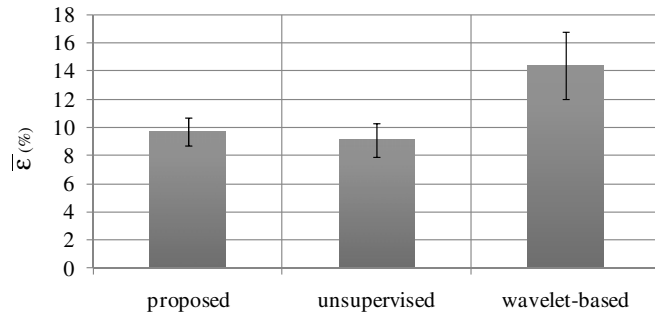
Figure 8. Relative error in the measurement of the extent of pulmonary infiltrates. The performance of the proposed methodology is compared with the unsupervised [12] and the wavelet-based methodologies [11]. The error represents two standard deviation units.
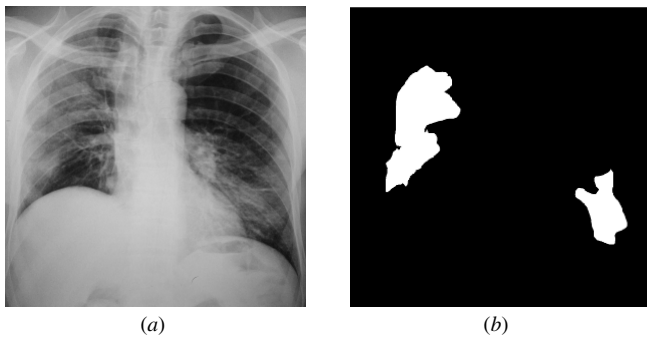

(*a*)   (*b*)

Figure 9. Ground truth areas of infiltrates as they have been annotated by experts

classified as ambiguous in the first step. Thus, for $T < 0.8$ the predicted class of a sample depends more on its textural properties than on the distribution of its intensity values and as such its classification is not possible in only the first step. Moreover, by conveying more samples in the second step, not only the variance introduced in the final result rises, but also the processing time increases. The latter is a consequence of the higher computational complexity involved in the calculation of the textural features in contrast to the calculation of the intensity distribution signatures.

On the other hand, for thresholds less than 0.8 a sharp decrease in the CE values is observed. This can be attributed to the fact that for the specific threshold range, many samples, that otherwise would have been classified as ambiguous, are incorrectly classified from the first step as infiltrates or as normal ones. This leads to an increase in the number of false positives (fp) and false negatives (fn) that is subsequently reflected in the CE rates. In contrast to the proposed classification method, both the unsupervised scheme and the wavelet-based scheme demonstrated inferior performance, yielding error rates of $33.2\% \pm 2.1\%$ and $28.7\% \pm 3.7\%$, respectively. It should be noted that if a chest radiograph is normal, without any sign of infiltrates, then the unsupervised scheme will always produce a false positive result because it should split the data into two target clusters. As a result the use of the unsupervised scheme is restricted only to radiographs that demonstrate infiltrates.

In order to assess the measurement capabilities of the proposed scheme a series of experiments was performed

that followed the leave-one-radiograph-out cross-validation process described earlier. To this end, the measurement error was employed so as to evaluate how effectively the proposed methodology measures the relative extent of the infection (R) in a radiograph. Measurement error was calculated by considering the threshold value that yielded the lowest CE rate ($T = 0.8$).

Since the radiographs under test may comprise both radiographs that may manifest infiltrates as well as normal radiographs, the error is measured in terms of $N = (1 - R)$, i.e. the ratio of normal pixels to the total number of pixels, so as to be properly quantified. Thus error is defined here as the relative error [31] between the true value of N as was derived from the ground truth ($N_{\text{true}} = 1 - R_{\text{true}}$) and the value of N approximated by the proposed methodology ($N_{\text{approx}} = 1 - R_{\text{approx}}$):

$$\varepsilon = \frac{|N_{\text{true}} - N_{\text{approx}}|}{N_{\text{true}}} \cdot 100\%. \qquad (10)$$

This calculation was repeated for every radiograph in the dataset and finally, the total mean measurement error and the total standard deviation were used so as to validate the proposed measurement methodology. Figure 8 illustrates a diagram with the estimated error rates obtained with the proposed methodology, the unsupervised scheme proposed in [12] and the wavelet-based methodology proposed in [11]. As can be observed, the proposed classification method results in comparable results to the one achieved by the unsupervised methodology while it outperforms the wavelet-based method, yielding error rates of $9.7\% \pm 1.0\%$, $9.1\% \pm 1.2\%$ and $14.4\% \pm 2.4\%$, respectively.

The present results were obtained on a set of chest radiographs exhibiting infiltrates in at least one of the two lungs so as to assess the measurement capabilities of all three schemes. In the proposed supervised classification approach the number of classes is predefined; however, it has the advantage of classifying the image patterns based on prior knowledge, and according to this knowledge it is allowable for a class to have zero members. Therefore, unlike the unsupervised approach, in the case of a chest radiograph without any sign of infiltrates, the supervised one is capable of returning a zero measurement for the infection's extent.

An illustrative example of the results obtained by the application of the proposed supervised framework versus the unsupervised methodology in [12] and the wavelet-based approach in [11], applied in the radiograph of figure 9 is presented in figure 10. The radiograph in figure 9 demonstrates infiltrates. The error rates obtained by equation (10) are 8.1% for the proposed scheme, 9.7% for the unsupervised one and 19.4% for the wavelet-based approach. In general, it is observed that the accuracies obtained by the proposed scheme outperform those obtained by the wavelet-based approach. Moreover, the proposed scheme demonstrates results that are better or close to those obtained by the unsupervised scheme, whereas the proposed one is better in classifying the infiltrates correctly. This derives from the fact that the areas identified as infiltrates are much closer to the ground truth than those identified by the unsupervised methodology. An example of normal chest radiograph where the unsupervised scheme
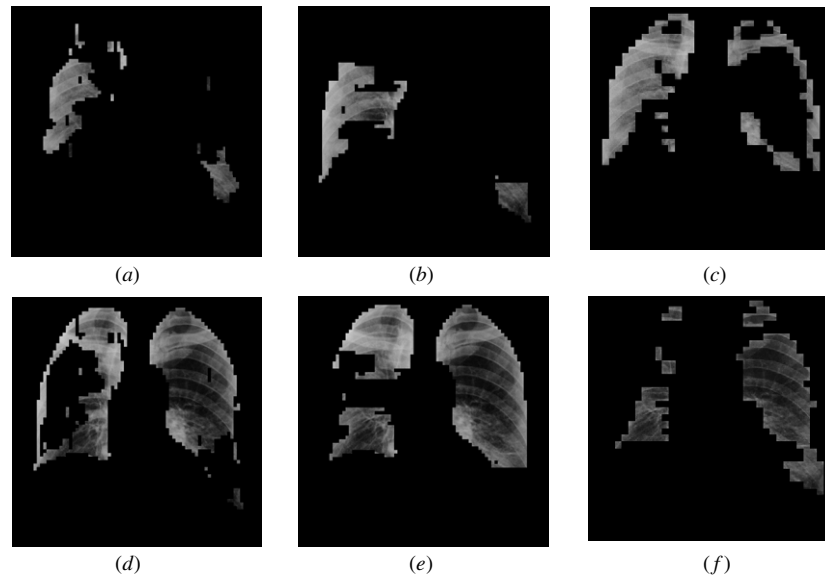
Figure 10. Images resulting from the assessment of figure 8(*a*) by the proposed supervised hierarchical classification scheme (left column), by the unsupervised scheme [12] (middle column) and by the wavelet-based approach [11] (right column). (*a*) Infiltrates as identified by the proposed scheme. (*b*) Infiltrates as identified by the unsupervised scheme. (*c*) Infiltrates as identified by the wavelet-based approach. (*d*) Normal lung parenchyma as identified by the proposed scheme. (*e*) Normal lung parenchyma as identified by the unsupervised scheme. (*f*) Normal lung parenchyma as identified by the wavelet-based approach.
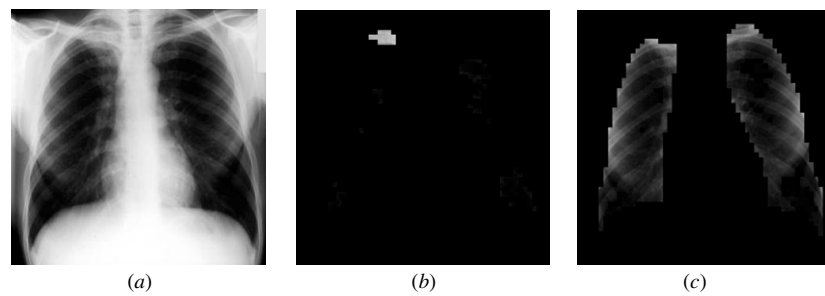

Figure 11. Application of the unsupervised method on a normal chest radiograph. (*a*) Original image, (*b*) false detection of an infiltrate with the unsupervised scheme proposed in [12], resulting in a false measurement of the relative extent of the infection, (*c*) normal lung parenchyma as identified by the unsupervised scheme.

results in a false measurement (error 6%) is illustrated in figure 11. In this case the proposed method returns zero infiltrates, i.e. the error is 0%.

4. Conclusions

We presented a novel pattern recognition framework enabling the assessment of the relative extent of pulmonary infiltrates in chest radiographs. This framework involves the extraction of patient-specific features suitable for the representation of radiographic intensity and texture across a diverse set of chest radiographs. The extracted features form patterns which are classified according to a supervised hierarchical classification scheme extending the one proposed in [12].

The conclusions that can be derived from this study are summarized as follows.

- The introduction of patient-specific information into the feature extraction process through image normalization enhances the analysis of chest radiographs.

- The proposed normalization approach results in intensity patterns of normal lung parenchyma and pulmonary infiltrates that exhibit high similarity across diverse sets of chest radiographs.

- The accuracy of the proposed supervised hierarchical classification scheme in the case of chest radiographs with pulmonary infiltrates is comparable to the one obtained by its predecessor classification approach which was unsupervised [12]. However, its advantage over the latter is that it has the ability to distinguish whether infiltrates exist or not.

- The proposed framework has been applied on chest radiographs; however, it could be applicable to a variety of medical imaging modalities. For example, the patient-specific normalization could be achieved using another characteristic region of normal tissue instead of the mediastinum/spinal cord region as a reference. There are also other imaging modalities where grey-level intensity and texture are significant features that could be

hierarchically considered, as for example in ultrasound imaging [25].

Future perspectives include the application of the proposed methodology in a decision support framework that will serve as a second opinion tool to support physicians in daily clinical practice, and integration into a multimodal data mining system for related adverse event detection.

Acknowledgments

References

[1] Iakovidis D K, Savelonas M A and Papamichalis G 2009 Robust model-based detection of the lung field boundaries in portable chest radiographs supported by selective thresholding *Meas. Sci. Technol.* **20** 104019

[2] Zhang J, Yan C-H, Chui C-K and Ong S H 2010 Accurate measurement of bone mineral density using clinical CT imaging with single energy beam spectral intensity correction *IEEE Trans. Med. Imaging* **29** 1382–9

[3] Arigovindan M, Suhling M, Jansen C, Hunziker P and Unser M 2007 Full motion and flow field recovery from echo Doppler data *IEEE Trans. Med. Imaging* **26** 31–45

[4] Müller N L, Franquet T, Lee K S, Isabela C and Silva S 2006 *Imaging of Pulmonary Infections* (Philadelphia, PA: Williams & Wilkins)

[5] Smith D L, Dushof J, Perencevich E N, Harris A D and Levin S A 2004 Persistent colonization and the spread of antibiotic resistance in nosocomial pathogens: resistance is a regional problem *Proc. Natl Acad. Sci. USA* **101** 3709–14

[6] Ginneken B V, Romeny B T H and Viergever M A 2001 Computer-aided diagnosis in chest radiography: a survey *IEEE Trans. Med. Imaging* **20** 1228–41

[7] Loog M and van Ginneken B 2006 Segmentation of the posterior ribs in chest radiographs using iterated contextual pixel classification *IEEE Trans. Med. Imaging* **25** 602–11

[8] Coppini G, Diciotti S, Falchini M, Villari N and Valli G 2003 Neural networks for computer-aided diagnosis: detection of lung nodules in chest radiograms *IEEE Trans. Inf. Technol. Biomed.* **7** 344–57

[9] Ginneken B V, Katsuragawa S, Romeny B T H, Doi K and Viergever M A 2002 Automatic detection of abnormalities in chest radiographs using local texture analysis *IEEE Trans. Med. Imaging* **21** 139–49

[10] Li X X, Wan S and Gong Y 2006 Mining x-ray images of SARS patients *Data Mining (Lecture Notes in Computer Science* vol 3755*)* ed G J Williams and S J Simoff (Berlin: Springer) pp 282–94

[11] Oliveira L L G, Silva S A, Vilela Ribeiro L H, Maurício de Oliveira R, Coelho C J and Andrade A L S S 2008 Computer-aided diagnosis in chest radiography for detection of childhood pneumonia *Int. J. Med. Inform.* **77** 555–64

[12] Tsevas S, Iakovidis D K and Papamichalis G 2010 *Engineering Intelligent Systems, Special Issue on Artificial Intelligence Applications and Innovations* vol 17 (Leicester: CRL Publishing) no 2/3

[13] Iakovidis D K, Savelonas M A and Papamichalis G 2009 Robust model-based detection of the lung field boundaries in portable chest radiographs supported by selective thresholding *Meas. Sci. Technol.* **20** 104019

[14] Iakovidis D K and Papamichalis G 2008 Automatic segmentation of the lung fields in portable chest radiographs based on Bézier interpolation of salient control points *Proc. IEEE Int. Conf. on Imaging Systems and Techniques (Chania, Greece)* pp 82–7

[15] Tsevas S and Iakovidis D K 2010 Patient specific normalization of chest radiographs and hierarchical classification of bacterial infection patterns *Proc. IEEE Int. Conf. on Imaging Systems and Techniques (IST) (Thessaloniki, Greece)* pp 156–60

[16] Shiraishi J, Katsuragawa S, Ikezoe J, Matsumoto T, Kobayashi T, Komatsu K, Matsui M, Fujita H, Kodera Y and Doi K 2000 Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists detection of pulmonary nodules *Am. J. Roentgenol.* **174** 71–4

[17] Aujol J F and Chan T F 2006 Combining geometrical and textured information to perform image classification *J. Vis. Commun. Image Represent.* **17** 1004-23

[18] Grigorescu S E, Petkov N and Kruizinga P 2002 Comparison of texture features based on Gabor filters *IEEE Trans. Image Process.* **11** 1160–7

[19] Jain A K and Farrokhnia F 1991 Unsupervised texture segmentation using Gabor filters *Pattern Recognit.* **24** 1167–86

[20] Fogel I and Sagi D 1989 Gabor filters as texture discriminator *Biol. Cybern.* **61** 103–13

[21] Tan T N 1995 Texture edge detection by modeling visual cortical channels *Pattern Recognit.* **28** 1283–98

[22] Muneeswaran K, Ganesan L, Arumugam S, K R and Soundar 2006 Texture image segmentation using combined features from spatial and spectral distribution *Pattern Recognit. Lett.* **27** 755–64

[23] Petkov N 1995 Biologically motivated computationally intensive approaches to image pattern recognition *Future Gener. Comput. Syst.* **11** 451–65

[24] Spitzer H and Hochstein S 1985 A complex-cell receptive-field model *J. Neurosci.* **53** 1266–86

[25] Keramidas E, Maroulis D and Iakovidis D K 2010 TND: a thyroid nodule detection system for analysis of ultrasound images and videos accepted for publication *J. Med. Syst* doi:10.1007/s10916-010-9588-7

[26] Ripley B D 1994 Neural networks and related methods for classification *J. R. Stat. Soc.* B **56** 409–56

[27] Swain M J and Ballard D H 1991 Color indexing *Int. J. Comput. Vis.* **7** 11–32

[28] Welch B L 1947 The generalization of 'Student's' problem when several different population variances are involved *Biometrika* **34** 28–35

[29] Weiss G M and Provost F 2001 The effect of class distribution on classifier learning *Technical Report* ML-TR-43 Department of Computer Science, Rudgers University

[30] Han J and Kamber M 2001 *Data Mining: Concepts and Techniques* (San Francisco, CA: Morgan Kaufmann Publishers)

[31] Taylor J 1997 *An Introduction to Error Analysis* 2nd edn (Sausalito, CA: University Science Books)