Pneumonia Detection on Chest X-Ray Using Machine Learning Paradigm

Tej Bahadur Chandra and Kesari Verma

Abstract The chest radiograph is the globally accepted standard used for analysis of pulmonary diseases. This paper presents a method for automatic detection of pneumonia on segmented lungs using machine learning paradigm. The paper focuses on pixels in lungs segmented ROI (Region of Interest) that are more contributing toward pneumonia detection than the surrounding regions, thus the features of lungs segmented ROI confined area is extracted. The proposed method has been examined using five benchmarked classifiers named Multilayer Perceptron, Random forest, Sequential Minimal Optimization (SMO), Logistic Regression, and Classification via Regression. A dataset of a total of 412 chest X-ray images containing 206 normal and 206 pneumonic cases from the ChestX-ray14 dataset are used in experiments. The performance of the proposed method is compared with the traditional method using benchmarked classifiers. Experimental results demonstrate that the proposed method outperformed the existing method attaining a significantly higher accuracy of 95.63% with the Logistic Regression classifier and 95.39% with Multilayer Perceptron.

Keywords Chest X-Ray · Consolidation · Pneumonia · Radiography · Thoracic disease · Pulmonary disease · Segmentation

1 Introduction

Medical imaging has a significant role in the categorization/classification of diseases. Chest X-ray is a medical imaging technology that is economical and easy to use. It produces an image of the chest, lung, heart, and airways (trachea) [1]. White consolidation on chest X-ray is the infection caused by bacteria, viruses, fungi, parasites, etc., within the small air spaces of the lungs results in an inflammatory response which is caused due of various abnormalities like pneumonia, tuberculosis, pneumothorax,

T. B. Chandra (✉) · K. Verma
Department of Computer Applications, National Institute of Technology, Raipur, India
e-mail: tejbahadur1990@gmail.com

K. Verma
e-mail: kverma.mca@nitrr.ac.in

B. B. Chaudhuri et al. (eds.), *Proceedings of 3rd International Conference on Computer Vision and Image Processing*, Advances in Intelligent Systems and Computing 1022,
https://doi.org/10.1007/978-981-32-9088-4_3

pleural effusion, etc. These inflammations can be easily seen as white patches on the chest X-ray (CXR). The vague appearance and resemblance with many other pulmonary abnormalities make it challenging to diagnose pneumonia for both radiologists as well as for automated computer-aided diagnostic (CAD) systems [2]. The invention of machine learning algorithms has motivated the researchers to develop a fully automated system that can assist radiologist as well as serve the purpose in resource-constrained remote areas. This paper presents a method for automatic detection of pneumonia by using lungs segmented ROI confined feature extraction method on CXR images. Instead of using the whole CXR image for analysis, which may lead to false diagnosis, the method considered only the lungs air space cavity by lungs segmented ROI selection. This type of automated system can be of great use in remote rural areas where the number of patients is very large as compared to available experienced radiologists.

1.1 Related Work

Automated computer-aided diagnosis (CAD) is becoming popular day by day. In the past few years, various methods have been proposed to improve the accuracy of pulmonary disease detection on CXRs. A significant work on abnormality detection (especially tuberculosis) on CXR can be found in [3–6]. Antani [7] has presented valuable work in the field of automated screening of pulmonary diseases including tuberculosis on digital CXR images. Ahmad et al. [8] compared the performance of different classifiers for abnormality analysis in CXRs. Karargyris et al. [9] proposed a new approach that combines both texture and shape feature for more accurate detection of tuberculosis and pneumonia. In a step toward building a mass screening system with high precision diagnosis, Wang et al. [10] proposed a new dataset namely “ChestX-ray8” and demonstrated thoracic disease localization using the deep convolutional neural network. This work has been further improved by Yao et al. [11] by considering the interdependency among the thoracic diseases and by Rajpurkar et al. [12] using deep learning approach with the 121-layer convolutional neural network.

Automatic lung field segmentation has drawn considerable attention of the researchers. The noteworthy contribution by Ginneken et al. [13] suggests that lung segmentation shows a significant improvement in the accuracy of the abnormality detection. Further, the author has compared the active shape, active appearance, and multi-resolution pixel classification model and observed nearly equal performance in each. A fully automatic technique for suppressing the ribs and clavicle in the chest X-ray is described in [14]. Shi et al. [15] have proposed patient-specific shape statistics and population-based deformable model to segment lung fields. Annangi et al. [16] used prior shape and low-level features based on the active contour method for lung segmentation. The groundbreaking work has been proposed by Candemir et al. [13] for automatic lung segmentation using anatomical atlases with nonrigid deformable registration.

Various features proven to be significant for the analysis of medical radiographs include shape feature [9, 16], texture feature: Histogram of Oriented Gradients (HOG) [5, 7, 17], Local Binary Patterns (LBP) [5, 7, 17, 18], Gray-Level Co-Occurrence Matrix (GLCM) based features [6], and Gabor feature [7, 8, 19].

1.2 Organization

The rest of the paper is organized as follows. Section 2 presents the detailed description of the dataset used to train and test the model, lungs segmentation, feature extraction methods, and different classifiers used to evaluate the classification accuracy. Section 3 elaborates the experimental setup followed by results and discussion in Sect. 4, and future scope in Sect. 5.

2 Materials and Methods

2.1 Dataset

In this paper, we used 412 CXR images containing 206 normal and 206 pneumonic cases from NIH ChestX-ray14 dataset [10] to train and test the classifier. The dataset contains a total of 112,120 frontal-view chest X-rays images of 30,805 unique patients with the text-mined fourteen thoracic disease labels. The disease labels are expected to have accuracy greater than 90%. All the CXR images used in this study are in PNG (Portable Network Graphics) grayscale format and having a resolution of 1024 dpi. Further, the lung atlas images from the set of reference images provided with the segmentation algorithm [13] are used to train the lung segmentation model.

2.2 Segmentation of Lungs Region and Preprocessing

Extracting features from the segmented portion of the lungs reduces the probable chances of the false-positive results. The organs like mediastinum, aortic knuckle, the apex of the heart, and the left and right diaphragm in CXR significantly affect the classification accuracy. This paper uses nonrigid registration-driven robust lung segmentation method proposed by Candemir et al. [13] for lungs segmented ROI selection. The method is organized in three stages as shown in Fig. 1.

Stage 1. In the first stage, the top five X-ray images that are most alike to the patient's X-ray are retrieved from a set of reference X-ray images using content-based image retrieval (CBIR) technique. The CBIR method employs partial radon transform and Bhattacharyya shape similarity measure to handle little affine distortion

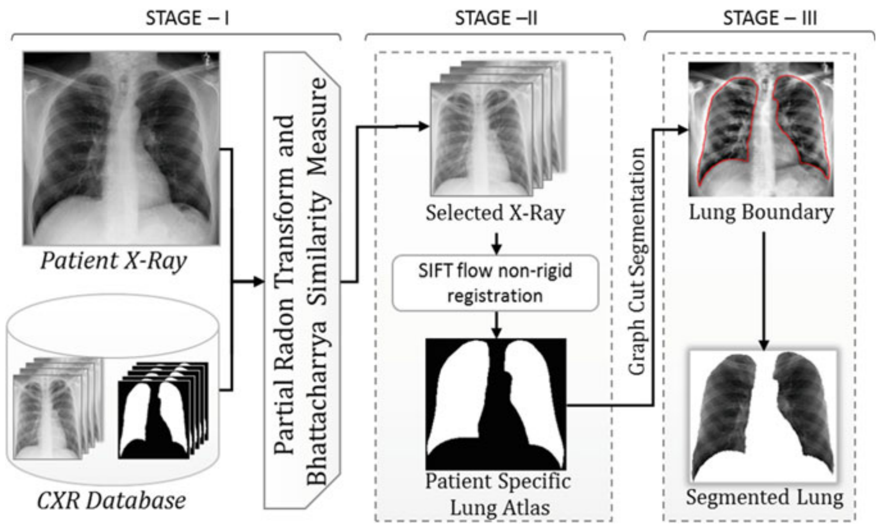

Fig. 1 Three stages of nonrigid registration-driven robust lung segmentation method (Candemir et al. 2014)

and to compute the degree of similarity between input and reference CXR images respectively.
Stage 2. In the second stage, the method employs SIFT flow deformable registration of the training mask to get the patient-specific modal of lung shape. The warped lung mask of top-ranked X-ray images is averaged to get the approximate lung modal.
Stage 3. Finally, in the third stage, the graph-cut method is used to extract the lungs boundary from patient-specific lung atlas using discrete optimization approach. The graph-cut method efficiently handles the problem of local minima using the energy minimization framework.

Subsequently, the binary mask extracted from the lung segmentation method is further used to cut the lungs region from the original CXR images.
ROI Selection: Image preprocessing is an important step in automated medical images analysis. The obtained segmented CXR images contain large white regions around the lungs segmented ROI which considerably affect the discriminating criteria of the features. Thus, in order to exclude it from feature extraction, the paper uses the lungs segmented ROI confined feature extraction technique which replaces all the pixels around the lungs segmented ROI with NaN (Not a Number) (as shown in Fig. 2).

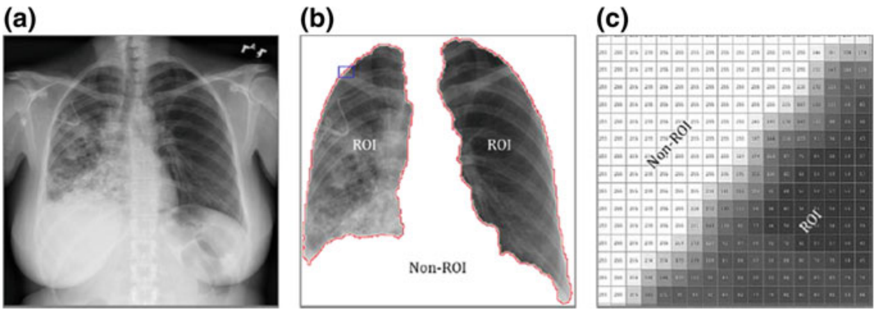

Fig. 2 **a** Full chest X-Ray image. **b** Segmented lungs showing bounded ROI Area. **c** Pixel values of blue patch showing ROI and non-ROI area

2.3 *Feature Extraction*

The presence of opaque consolidations on CXR due to pneumonia can be analyzed using statistical features. In this study, eight statistical features based on the first-order histogram of the original image [20] are used. The features extracted using this method does not take into account the neighboring pixel relationships. The features are summarized in Table 1. The *P(I)* represents the first-order histogram at the gray level *I* of an image having total *L* gray levels.

$$P(I) = \frac{\mathrm{N(I)}}{\mathrm{N}} = \frac{\text{number of pixels with gray level I}}{\text{total number of pixels in the region}}$$

Table 1 Eight first-order statistic features (Srinivasan et al. 2008)

Code	Feature	Equation
F1	Mean (m)	$\sum_{I=0}^{L-1} IP(I)$
F2	Variance (μ2)	$\sum_{I=0}^{L-1} (I-m)^2 P(I)$
F3	Standard deviation (σ)	$\sqrt{\mu_2}$
F4	Skewness (μ3)	$\sum_{I=0}^{L-1} (I-m)^3 P(I)$
F5	Kurtosis (μ4)	$\sum_{I=0}^{L-1} (I-m)^4 P(I)$
F6	Smoothness (R)	$1-\frac{1}{1+\sigma^2}$
F7	Uniformity (U)	$\sum_{I=0}^{L-1} P^2(I)$
F8	Entropy (e)	$\sum_{I=0}^{L-1} P(I)\log_2 P(I)$

2.4 Classification

The pneumonia detection performance of the proposed method is evaluated using five benchmarked classifiers namely Multilayer Perceptron (MLP), Random forest, Sequential Minimal Optimization (SMO), Classification via Regression, and Logistic Regression classifier available in Weka (Waikato Environment For Knowledge Analysis) tool, version 3.8.2 [21]. The classifiers' performance are evaluated with 10-fold cross-validation. Following are the short descriptions of each classifier:

a. *Multilayer Perceptron (MLP)*: It mimics the working of the human brain by utilizing backpropagation with adjustable weight w_{ij} for training its neurons. The trained model can solve very complex classification problems stochastically and can also deal with nonlinearly separable data [22, 23].
b. *Random Forest*: This supervised classification and regression method works by generating a forest (ensemble) of decision trees. It uses majority voting to decide the final class labels of the object. This classifier works efficiently on large databases and even handles missing data [24].
c. *Sequential Minimal Optimization (SMO)*: It is used to train the support vector machine (SVM) classifier by decomposing large quadratic programming problems into sub-problems. The multiclass problems are solved using pairwise classification [25].
d. *Classification via Regression*: It builds a polynomial classification model that uses regression method for classification through the given regression learner. For each binarized class labels, regression learner generates one regression model which predicts a value from a continuous set, whereas classification predicts the belongingness to a class [26].
e. *Logistic Regression*: It is the statistical method that uses a logistic sigmoidal function to compute probability value which can be mapped to two or more discrete classes. It is most widely used in medical fields for risk prediction based on observed characteristics of the patient [27].

3 Experimentation and Performance Evaluation

The proposed structural modal for machine learning based diagnosis of pneumonia is shown in Fig. 3. The modal works in two phases: training and testing. Both the phases include image preprocessing followed by feature extraction and classification. In the training phase, eight first-order statistical features are extracted from a selected set of 412 CXR images. The extracted features are now utilized to build training models using benchmarked classifiers whereas in the testing phase, the extracted features from the patient X-ray are fed to the classifier and the predicted class label is compared with the available ground truth data.

The performance of different benchmarked classifiers applied on the preprocessed dataset is evaluated and compared using six performance metrics, namely Accuracy,

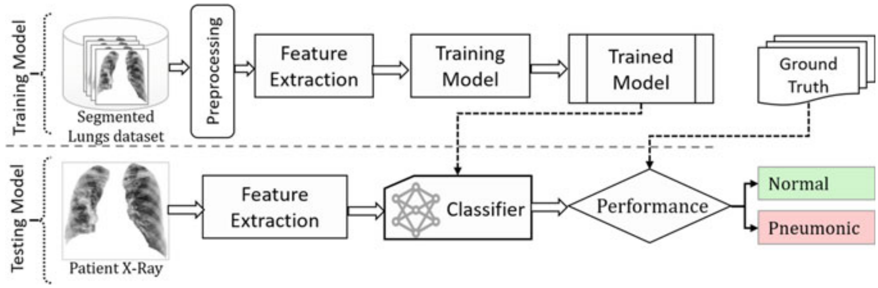

Fig. 3 Structural design of the experimental setup

Sensitivity or Recall, Specificity, Precision (PPV), area under the Receiver Operating Characteristic curve (AUROC), and F1 Score [22]. Short descriptions of the performance metrics are shown in Table 2 where TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative, P = TP + FN, and N = TN + FP.

Table 2 Short description of performance metrics [22]

Measure (in %)	Equation	Description
Accuracy	$\frac{TP+TN}{P+N} \times 100$	Percentage of correctly classified test cases by the classifier
Sensitivity/recall	$\frac{TP}{P} \times 100$	The proportion of pneumonia cases that are correctly identified to the total actual pneumonic cases. Also known as true-positive or recognition rate
Specificity	$\frac{TN}{N} \times 100$	The proportion of normal cases that are correctly identified to total actual normal cases. Also known as true-negative rate
Precision (PPV)	$\frac{TP}{TP+FP} \times 100$	The proportion of pneumonia cases that are correctly identified to total predicted pneumonic cases. Also known as positive predicted value
Area under curve (AUC)	$\frac{1}{2}\left(\frac{TP}{P} + \frac{TN}{N}\right)$	It is the common measure of the performance of the classifier using sensitivity and specificity
F1 score	$\frac{2 \times precision \times recall}{precision + recall}$	It is the harmonic mean of precision and recall and gives equal weight to both false-positive and false-negative cases. It is used to measure the test's accuracy

4 Results and Discussions

The automated computer-aided diagnosis of pneumonia using the whole chest X-ray image suffers from false-positive results due to the presence of various organs like heart, thymus gland, aortic knuckle, sternum, diaphragm, spine, etc., in thorax regions. The proposed method overcomes this limitation by using segmented lungs regions and ROI confined feature extraction. The experiments were conducted using a full chest X-Ray image and segmented lungs region are shown in Tables 3 and 4, respectively. From the experimental results, it can be observed that the pneumo-

Table 3 Classifiers' performance prior to lung segmentation (using full chest X-ray image)

Classifier	Performance measures					
	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	AUC	F1 score (%)
Multilayer Perceptron	**92.233**	**86.408**	**98.058**	**97.802**	**0.922**	**91.753**
Random forest	90.534	86.408	94.660	94.180	0.905	90.127
Sequential Minimal Optimization	89.806	80.097	99.515	99.398	0.898	88.710
Classification via Regression	91.990	86.408	97.573	97.268	0.920	91.517
Logistic Regression	91.505	86.408	96.602	96.216	0.915	91.049

Table 4 Classifiers performance after lung segmentation (using ROI confined feature extraction)

Classifier	Performance measures					
	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	AUC	F1 score (%)
Multilayer Perceptron	**95.388**	**93.204**	**97.573**	**97.462**	**0.954**	**95.285**
Random forest	94.417	93.689	95.146	95.074	0.944	94.377
Sequential Minimal Optimization	93.689	89.320	98.058	97.872	0.937	93.401
Classification via Regression	94.660	91.262	98.058	97.917	0.947	94.472
Logistic Regression	**95.631**	**93.689**	**97.573**	**97.475**	**0.956**	**95.545**

nia detection performance using segmented lungs regions surpasses the traditional method that uses full chest X-ray image. This is because, in traditional methods, the extracted features are significantly affected by Non-ROI portions leading to misclassification while the proposed method extracts feature from the segmented lung ROI that are contributing more toward disease classification. To assess and compare the performance of pneumonia detection, five benchmarked classifiers are used. From Table 3, it can be observed that Multilayer Perceptron outperforms the other benchmarked classifiers achieving higher accuracy of 92.23% when used with features extracted from full chest X-ray image (without lung segmentation).

However, the overall disease detection performance is improved when the segmented lungs region is used (as shown in Table 4) as the pixels in ROI are contributing more toward pneumonia detection than the surrounding regions.

The classification performance of the proposed method using segmented lungs region shown in Table 4 suggests that the Logistic Regression classifier and Multilayer Perceptron outperform the others benchmarked classifiers. The highest performance with an accuracy of 95.63% is obtained using Logistic Regression classifier whereas the Multilayer Perceptron also gives nearly equal performance with an accuracy of 95.39% under the same configuration. The average percentage increase in performance metrics of the benchmarked classifiers using the segmented regions is shown in Table 5. From the table, it is found that the lung segmentation based ROI feature extraction approach shows an average increase in the accuracy of each classifier by 3.54%. The significant upswing in performance is due to ROI confined feature extraction which considers only the pixels in ROI on segmented lungs as shown in Fig. 2b. The Logistic Regression classifier shows comparably noteworthy

Table 5 Percentage increase in classifier performance after lung segmentation (using ROI confined feature extraction)

Classifier	Performance measures					
	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	AUC	F1 score (%)
Multilayer Perceptron	3.155	6.796	−0.485	−0.340	0.032	3.533
Random forest	3.883	7.282	0.485	0.894	0.039	4.250
Sequential Minimal Optimization	3.883	9.223	−1.456	−1.525	0.039	4.691
Classification via Regression	2.670	4.854	0.485	0.649	0.027	2.956
Logistic Regression	4.126	7.282	0.971	1.259	0.041	4.496
Average increase	**3.544**	**7.087**	**0.000**	**0.187**	**0.035**	**3.985**

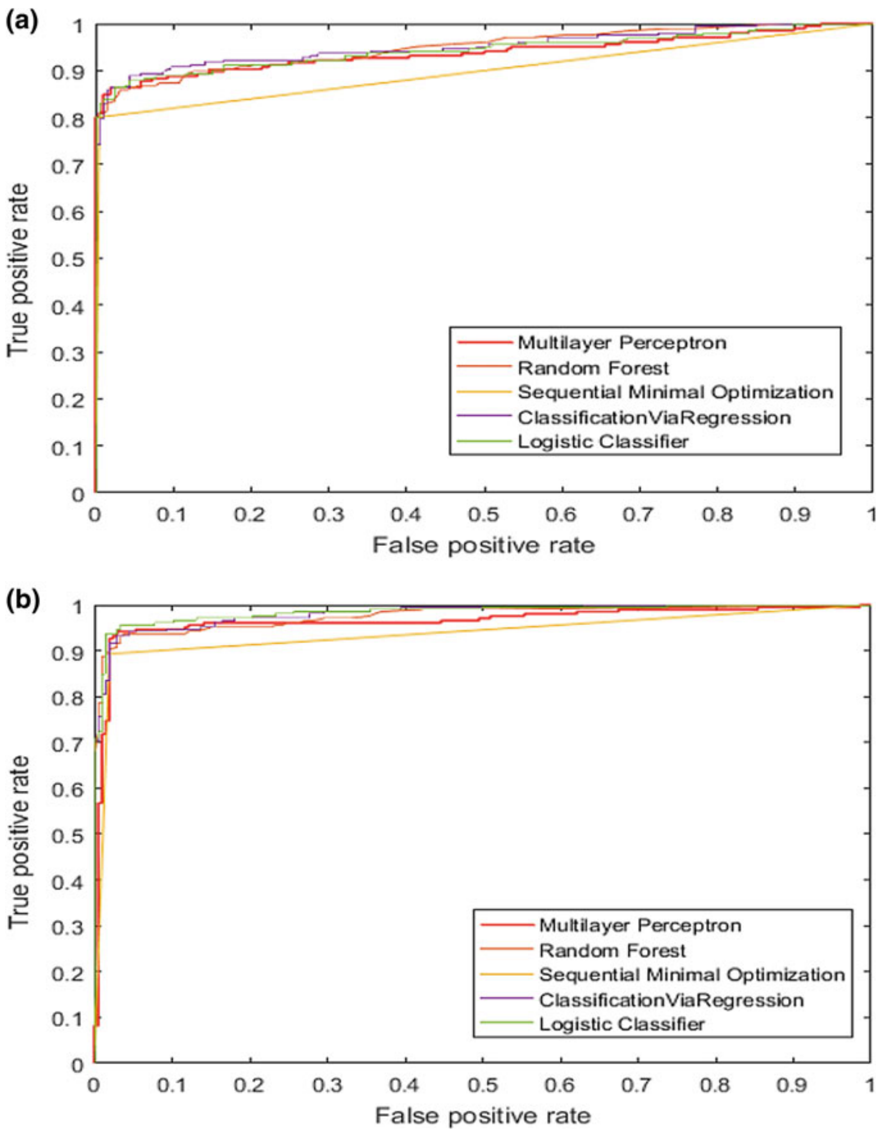

Fig. 4 **a** Area under ROC curve of five classifiers prior to lung segmentation (using full chest X-ray image). **b** The area under ROC curve of five classifiers after lung segmentation

improvement in performance than others with an increase in accuracy of 4.13%; furthermore, the sensitivity, specificity, precision, AUC, and F1 score are increased by 7.28%, 0.97%, 1.26%, 0.041, and 4.50%, respectively.

Performance metrics shown in Tables 3 and 4 provide us with measures to assess the classification performance with a fixed threshold (cutoff) value on class probability. However, different thresholds result in different accuracy, sensitivity, specificity, precision, and F1 score values. Thus, in order to measure the overall performance of the classifier model over its entire operating range, area under ROC (receiver operating characteristic) curve (AUROC) is used. The ROC analysis offers tools to make an optimal selection of classifier models discarding suboptimal ones independently from the cost context or the class distribution. The area under the curve (AUC) for the perfect classifier is 1.0 while the classifier with no power has the AUC of 0.5. The AUCs of most classifier fall between these two values. Figure 4a and b show the plots of the area under the ROC curve of five benchmarked classifiers using full chest X-ray image prior to lung segmentation and after lung segmentation, respectively. From the analysis of ROC plots, it is observed that AUC using the segmented lungs region is significantly higher than using full chest X-ray image. The AUC of the Logistic Regression classifier and Multilayer Perceptron with value 0.956 and 0.954, respectively, using segmented lungs regions fall very close to 1.0 achieving a higher rank among benchmarked classifier models.

5 Conclusion and Future Work

Automatic detection of pneumonia using chest radiograph is still challenging using CAD systems. This paper presents a method for automatic detection of pneumonia using the statistical feature of the lungs airspace. The method is based on the analysis of nonrigid deformable registration driven automatically segmented lungs regions and lungs segmented ROI confined feature extraction. Experiments performed on 412 chest X-ray images containing 206 normal and 206 pneumonic cases from ChestX-ray14 dataset suggest that the performance of the proposed method for automatic detection of pneumonia using segmented lungs is significantly better than traditional method using full chest X-ray images. The average accuracy of the proposed method is 3.54% higher than the traditional method. The Logistic Regression classifier with disease detection accuracy of 95.63% outperforms the other benchmarked classifiers using segmented lungs regions. However, despite high accuracy, the method needs more reliable feature analysis techniques and rigorous testing on thousands of real-time test cases.

The lung segmentation approach used in this paper needs further enhancement for pediatric cases and pleural effusion. The method completely ignores the clinical background of the patients; the integration of this method is a further aspect of research.

Acknowledgements The authors would like to thank Dr. Javahar Agrawal, Diabetologist and Senior Consulting Physician, Lifeworth Super Speciality Hospital, Raipur and Dr. A. D. Raje, Consulting Radiologist, MRI Diagnostic Institute, Choubey Colony, Raipur for their valuable guidance.

References

1. Dong, Y., Pan, Y., Zhang, J., Xu, W.: Learning to read chest X-Ray images from 16000+ examples using CNN. In: Proceedings of the Second IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies
2. Van Ginneken, B., Ter Haar Romeny, B.M., Viergever, M.A.: Computer-aided diagnosis in chest radiography: a survey. IEEE Trans. Med. Imaging **20**, 1228–1241 (2001)
3. Mohd Rijal, O., Ebrahimian, H., Noor, N.M.: Determining features for discriminating PTB and normal lungs using phase congruency model. In: Proceedings—IEEE-EMBS International Conference on Biomedical and Health Informatics: Global Grand Challenge of Health Informatics, BHI 2012, vol. 25, pp. 341–344 (2012)
4. Van Ginneken, B., Philipsen, R.H.H.M., Hogeweg, L., Maduskar, P., Melendez, J.C., Sánchez, C.I., Maane, R., dei Alorse, B., D'Alessandro, U., Adetifa, I.M.O.: Automated scoring of chest radiographs for tuberculosis prevalence surveys: a combined approach. In: Fifth International Workshop on Pulmonary Image Analysis, pp. 9–19 (2013)
5. Jaeger, S., Karargyris, A., Candemir, S., Folio, L., Siegelman, J., Callaghan, F., Xue, Z., Palaniappan, K., Singh, R.K., Antani, S., Thoma, G., Wang, Y., Lu, P., Mcdonald, C.J.: Automatic tuberculosis screening using chest radiographs. Stefan **33**, 233–245 (2014)
6. V, R.D.: Efficient automatic oriented lung boundary detection and screening of tuberculosis using chest radiographs. J. Netw. Commun. Emerg. Technol. **2**, 1–5 (2015)
7. Antani, S.: Automated detection of lung diseases in chest X-Rays. US Natl. Libr. Med. (2015)
8. Ahmad, W.S.H.M.W., Logeswaran, R., Fauzi, M.F.A., Zaki, W.M.D.W.: Effects of different classifiers in detecting infectious regions in chest radiographs. In: IEEE International Conference on Industrial Engineering and Engineering Management 2015–January, pp. 541–545 (2014)
9. Karargyris, A., Siegelman, J., Tzortzis, D., Jaeger, S., Candemir, S., Xue, Z., Santosh, K.C., Vajda, S., Antani, S., Folio, L., Thoma, G.R.: Combination of texture and shape features to detect pulmonary abnormalities in digital chest X-rays. Int. J. Comput. Assist. Radiol. Surg. **11**, 99–106 (2016)
10. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R.M.: ChestX-ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases, pp. 2097–2106 (2017)
11. Yao, L., Poblenz, E., Dagunts, D., Covington, B., Bernard, D., Lyman, K.: Learning to diagnose from scratch by exploiting dependencies among labels, pp. 1–12 (2017). arXiv preprint arXiv:1710.10501
12. Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K., Lungren, M.P., Ng, A.Y.: CheXNet: radiologist-level pneumonia detection on chest X-rays with deep learning, pp. 3–9 (2017). arXiv preprint arXiv:1711.05225
13. Candemir, S., Jaeger, S., Palaniappan, K., Musco, J.P., Singh, R.K., Xue, Z., Karargyris, A., Antani, S., Thoma, G., McDonald, C.J.: Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. IEEE Trans. Med. Imaging **33**, 577–590 (2014)
14. Suzuki, K., Abe, H., MacMahon, H., Doi, K.: Image-processing technique for suppressing ribs in chest radiographs by means of massive training artificial neural network (MTANN). IEEE Trans. Med. Imaging **25**, 406–416 (2006)
15. Shi, Y., Qi, F., Xue, Z., Chen, L., Ito, K., Matsuo, H., Shen, D.: Segmenting lung fields in serial chest radiographs using both population-based and patient-specific shape statistics. IEEE Trans. Med. Imaging **27**, 481–494 (2008)
16. Annangi, P., Thiruvenkadam, S., Raja, A., Xu, H., Sun, X.S.X., Mao, L.M.L.: A region based active contour method for x-ray lung segmentation using prior shape and low level features. 2010 IEEE International Symposium on Biomedical Imaging From Nano to Macro, pp. 892–895 (2010)
17. Surya, S.J., Lakshmanan, S., Stalin, J.L.A.: Automatic tuberculosis detection using chest radiographs using its features abnormality analysis. J. Recent Res. Eng. Technol. **4** (2017)

18. Fatima, S., Irtiza, S., Shah, A.: A review of automated screening for tuberculosis of chest Xray and microscopy images. Int. J. Sci. Eng. Res. **8**, 405–418 (2017)
19. Scholar, P.G.: A robust automated lung segmentation system for chest X-ray (CXR) images. Int. J. Eng. Res. Technol. **6**, 1021–1025 (2017)
20. Srinivasan, G., Shobha, G.: Statistical texture analysis. In: Proceedings of World Academy of Science, Engineering and Technology, vol. 36, pp. 1264–1269 (2008)
21. Frank, E., Hall, M.A., Witten, I.H.: The WEKA workbench, 4th edn, pp. 553–571. Morgan Kaufmann (2016)
22. Han, J., Kamber, M., Pei, J.: Data mining: concepts and techniques (2012)
23. Haykin, S.: Neural networks: a comprehensive foundation. Prentice Hall (1998)
24. Breiman, L.: Random forests. Mach. Learn. **45**, 5–32 (2001)
25. Platt, J.C.: Sequential minimal optimization: a fast algorithm for training support vector machines. Adv. Kernel Methods, 185–208 (1998)
26. Frank, E., Wang, Y., Inglis, S., Holmes, G., Witten, I.H.: Using model trees for classification. Mach. Learn. **32**, 63–76 (1998)
27. Sperandei, S.: Understanding logistic regression analysis. Biochem. Medica. **24**, 12–18 (2014)