Potential of digital chest radiography-based deep learning in screening and diagnosing pneumoconiosis

Yajuan Zhang  
12th people's Hospital of Guangzhou

Bowen Zheng  
Nan fang Hospital, Southern Medical University

Long Li  
12th people's Hospital of Guangzhou

Fengxia Zeng  
Nan fang Hospital, Southern Medical University

Tianqiong Wu  
12th people's Hospital of Guangzhou

Xiaoke Cheng  
12th people's Hospital of Guangzhou

Yuli Peng  
12th people's Hospital of Guangzhou

Yonliang Zhang  
12th people's Hospital of Guangzhou

Yuanlin Xie  
San shui District Institute for Disease Control and Prevention

Wei Yi  
The Third People's Hospital of Yunnan Province  Yunnan  650010

Weiguo Chen  
Nan fang Hospital, Southern Medical University

Genggeng Qin  (zealotq@smu.edu.cn)  
Nan fang Hospital, Southern Medical University

Jiefang Wu  
Nan fang Hospital, Southern Medical University

Research Article

Keywords: Pneumoconiosis, Mass screening and classification, Radiography, Deep learning, Diagnosis

Posted Date: June 7th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2990485/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.  Read Full License
Abstract

Background

To improve the accuracy of pneumoconiosis diagnosis, a computer-assisted method was developed.

Methods

Three CNNs (Resnet50, Resnet101, and DenseNet) were used for pneumoconiosis classification based on 1,250 chest X-ray images. Three double-blinded experienced and highly qualified physicians read the collected digital radiography images and classified them from category 0 to category III. The results of the three physicians in agreement were considered the relative gold standards. Subsequently, three CNNs were used to train and test these images and their performance was evaluated using multi-class classification metrics. We used kappa values and accuracy to evaluate the consistency and reliability of the optimal model with clinical typing.

Results

ResNet101 was the optimal model among the three CNNs. The AUC of ResNet101 was 1.0, 0.9, 0.89, and 0.94 for detecting pneumoconiosis categories 0, I, II, and III, respectively. The micro-average and macro-average mean AUC values were 0.93 and 0.94, respectively. The accuracy and Kappa values of ResNet101 were 0.72 and 0.7111 for quadruple classification and 0.98 and 0.955 for dichotomous classification, respectively, compared with the relative standard classification of the clinic.

Conclusion

The ResNet101 model performed relatively better in classifying pneumoconiosis than radiologists. The dichotomous classification displayed outstanding performance, thereby indicating the feasibility of deep learning techniques in pneumoconiosis screening.

INTRODUCTION

Pneumoconiosis is a chronic occupational lung disease caused by the inhalation of productive mineral dust. It is incurable and irreversible, and is the leading occupational disease in China [1]. Chronic silicosis may develop or progress even after the cessation of occupational exposure; currently, there is no treatment other than a potential lung transplant [2–4]. This necessitates diagnosing and classifying the stage of pneumoconiosis before its progress into an irreversible stage. To reduce complication rates and mortality, the International Labor Organization (ILO) recommends frequent pulmonary function tests and chest radiographs for people with occupational diseases [5]. Researchers have developed a standardized system for classifying imaging abnormalities in pneumoconiosis according to the presence of the following pulmonary parenchymal and pleural abnormalities: small round turbidity, small irregular turbidity, massive turbidity, and other imaging features [6–8]. X-ray imaging is the most common modality used in clinical settings worldwide [9].

Currently, the clinical diagnosis of pneumoconiosis is principally based on the corrected interpretation of chest radiographs (X-ray images). Compared with standard radiographs, a radiologist assesses the concentration of small opacities on a chest X-ray image as category 0, I, II, or III [10]. While classifying pneumoconiosis, we followed the ILO classification guidelines; nonetheless, we used standard radiographs collected and defined by the Chinese Center for Disease Control and Prevention [11–12]. However, the diagnosis of pneumoconiosis remains challenging, is subjective, and varies among
reviewers [13–14]. To improve the diagnostic efficiency and accuracy among radiologists, researchers have developed computer-aided diagnosis (CAD) schemes for detecting pneumoconiosis using chest radiographs as a second opinion.

Several domestic and international scholars have studied the application of CAD technology in pneumoconiosis diagnosis [15–16]. However, these traditional machine learning methods rely on the effectiveness of feature extraction and require "hand-crafted" feature recognition, which is technically time-consuming and labor-intensive, particularly for complex tasks, such as pneumoconiosis diagnosis and staging [17].

The deep learning technique has emerged as a novel and promising approach for solving challenging problems. Its advantage is that it implicitly learns complex imaging features or patterns without recognizing and extracting image features, which may involve tens of millions of features and analyze them to obtain high-level features [18–19]. The deep learning network is an advanced technique for studying the aforementioned features and can offer heat maps that visualize the interpretation of model decisions. Thus, an improvement of the network reliability and an accurate representation of the content of the input image data solves the problem of the “black box” principle [20]. By comparing the performance of three convolutional neural networks (CNNs), we aimed to establish an end-to-end automatic pneumoconiosis medical image classification model based on deep learning and evaluate the feasibility of the model for pneumoconiosis classification. We also intended to conduct a comparative analysis of the assessment of this model with clinical assessment and evaluate the consistency among radiologists to determine the potential of this tool in practice.

**MATERIALS AND METHODS**

**Study population**

All private information was de-identified. All participants were industrial workers with a history of DR screening of dust exposure for pneumoconiosis from 2016 to 2019. Of these participants, 932 were diagnosed with pneumoconiosis and 318 were healthy. The study comprised 124 women and 1,126 men, including 23 women and 295 men in the healthy category. Their experience of dust service ranged from 1–40, with a mean duration of 20.56 ± 2.5 years. We did not exclude patients with emphysema, tuberculosis, bronchiectasis, or other structural lung diseases. This is because patients with pneumoconiosis may simultaneously experience these diseases.

**Data acquisition**

We retrospectively collected 1,250 cases of different stages of pneumoconiosis from three hospitals. The patients’ details are summarized in Table 1.
<table>
<thead>
<tr>
<th>Dataset origin</th>
<th>No. pneumonia</th>
<th>No. normal</th>
<th>File type</th>
<th>Women</th>
<th>men</th>
<th>Category 0</th>
<th>Category I</th>
<th>Category II</th>
<th>Category III</th>
</tr>
</thead>
<tbody>
<tr>
<td>Institution 1</td>
<td>569</td>
<td>209</td>
<td>DICOM</td>
<td>35</td>
<td>743</td>
<td>209</td>
<td>190</td>
<td>187</td>
<td>192</td>
</tr>
<tr>
<td>Institution 2</td>
<td>234</td>
<td>59</td>
<td>DICOM</td>
<td>56</td>
<td>237</td>
<td>59</td>
<td>92</td>
<td>70</td>
<td>72</td>
</tr>
<tr>
<td>Institution 3</td>
<td>129</td>
<td>50</td>
<td>DICOM</td>
<td>33</td>
<td>146</td>
<td>50</td>
<td>40</td>
<td>42</td>
<td>47</td>
</tr>
</tbody>
</table>

There were a total of 1,250 cases (932 cases positive for pneumoconiosis and 318 healthy controls). "Positive cases" refer to cases that were positive for pneumoconiosis.

DICOM = Digital Communications in Medicine.

All images are presented in digital radiography.

Images were acquired using Siemens Axiom Aristors and Definium 6000 X-ray radiographers (General Motors). The digital radiography (DR) image settings were as follows: 120 kVp and automated mAs; 120 kVp and 250 mAs. The final digital images were generated using a DR workstation. Image processing techniques, such as edge enhancement and noise reduction, were turned off in the post-processing software. These images were calibrated to comply with the Digital Imaging and Communications in Medicine standard.

**Annotation**

Following the ILO guidelines, three board-certified radiologists reviewed the images and classified them into four categories as follows: 0 (n = 318), I (n = 322), II (n = 299), and III (n = 311); during the diagnosis, the profusion, shape, and size of pneumoconiosis lesions were considered the most important evidence for classifying pneumoconiosis [21–22]. All radiologists were qualified for diagnosing pneumoconiosis, with > 10 years of experience. In addition, one radiologist participated in the formulation of the latest diagnostic standard for pneumoconiosis in China, the "Diagnostic Standard for Occupational Pneumoconiosis GBZ70-2015" [23]. We followed the strategy used by other researchers and divided the patients into two categories, namely normal and pneumoconiosis, by combining the I, II, and III categories into one category [24–25]. The absence of disagreement on the classification of an image required multiple rounds of discussion and adjudication until complete agreement was reached [26]. The categories of pneumoconiosis are provided in Appendix E1.

**Deep learning models**

Before training the CNNs, all images were processed, including down-sampling, histogram equalization, blank area removal, and down-sampling. Post processing, the images were resized to a 256×256 pixel matrix, converted to a JPG format, and inputted to the server to build the dataset (Appendix E2). In this study, deep learning models were built using three CNN architectures as follows: ResNet50, ResNet101, and DenseNet.

**CNNs training and validation**

Before CNN training, we randomly split all data into a training set (80%) and test set (20%). Remaining 20% of the images from the training set were extracted as the validation set.

During the training, the parameters were continuously adjusted followed by adding the validation set. We adjusted the network structure and training parameters until the training accuracy of the model and the number of training sessions were maximized. In the experimental process, categorical-cross-entropy was used as the loss function, whereas rectified linear activation unit was used as the activation function to adjust the network parameters and fit the training data.
Indicators, such as accuracy curves and loss rate curves, were monitored to observe changes in each parameter and to understand the training situation. Post training, we evaluated the effectiveness of the model by relevant evaluation indexes using the test set. The modeling strategy also mimicked the clinical diagnostic procedures, thus allowing us to merge pneumoconiosis-related knowledge into the artificial intelligence model. In addition, we requested three radiologists to independently read the images in the test dataset and compare the clinical diagnoses with that of the deep learning approach.

**General scheme of the four pneumoconiosis classifications**

Figure 1 depicts the general scheme of classification using a quadruple classification method. The network model was trained and supervised using the known labeled pneumoconiosis images. The backpropagation algorithm continuously adjusted the parameters of the network model to achieve the function of accurately classifying unknown images into four categories of DR images, that is, categories 0, I, II, and III.

**Evaluation metrics**

The accuracy, precision, and recall were used to evaluate the performance of the model. The evaluation indicators are presented in Appendix E3.

A cross-validation approach was used to evaluate the performance of the proposed deep learning model, and four expected results for true positive, true negative, false positive, and false negative were obtained.

A multi-categorization task necessitates the use of macro-P, macro-R, Macro-F1 and Macro-averages evaluation indexes. Data used in this study were divided into four categories. The four-category problem was transformed into four binary problems, followed by the calculation of the check-all rate of the confusion matrices. The evaluation indicators are listed in Appendix E3.

**Statistical analyses**

We used the ROC analysis and AUC to measure the diagnostic effectiveness of the classifier. We conducted a non-parametric ROC [27–28] analysis on an independent test dataset to evaluate the performance of the predictive models. Each point on the ROC curve represented a sensitivity/specificity pair corresponding to a specific decision threshold. In addition, we invited three radiologists (R1, R2, and R3) with > 10 years of experience in pneumoconiosis diagnosis and familiarity with pneumoconiosis diagnostic criteria. These certified radiologists independently read the images in the test dataset in a double-blinded manner and classified them into categories 0, I, II, and III.

We evaluated the performance of the reader using the ROC analysis and compared it with that of the deep learning algorithm. Moreover, we conducted statistical analyses to evaluate the consistency between model classification and clinical evaluation using Kappa coefficients. We determined the model classification accuracy using IBM SPSS V.20 software.

**RESULTS**

**Experimental results**

As observed from the curves, the accuracy rate increased and the loss rate decreased in the validation and training sets until the network converged with the highest accuracy rate and the lowest loss rate(Fig. 2).

Figure 3-a to 3-c and Table 2 present the classification results of each dataset for patients with pneumoconiosis in ResNet50, ResNet101, and DenseNet CNNs.
To evaluate the CNN performances, we plotted ROC curves. Figure 4 depicts the ROC curves of the three CNNs, which were subsequently summarized. The trained ResNet101 model demonstrated the best performance. The AUC values of ResNet101 micro-average and macro-average were 0.93 and 0.94, respectively, with AUC values of 1.0, 0.90, 0.89, and 0.94 for category 0, category I, category II, and category III, respectively.

**Diagnostic performance compared with that of radiologists**

The inconsistency rate between the model and the clinical assessment of pneumoconiosis classification was principally distributed in assessments for categories I/III and II/III, whereas the rate was lower for 0/I and other categories (0.016 and 0.033, respectively). Table 3 provide the distribution of inconsistent rates between ResNet101 and clinical assessment.

<table>
<thead>
<tr>
<th>Absolute inconsistency rate</th>
<th>0.72(86/120)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0/ Inconsistency rate</td>
<td>0.016(2/120)</td>
</tr>
<tr>
<td>/ Inconsistency rate</td>
<td>0.083(10/120)</td>
</tr>
<tr>
<td>/ Inconsistency rate</td>
<td>0.083(10/120)</td>
</tr>
<tr>
<td>two-degree Inconsistency rate</td>
<td>0.033(4/120)</td>
</tr>
<tr>
<td>Total</td>
<td>1(120/120)</td>
</tr>
</tbody>
</table>

The model assessed pneumoconiosis classification with high accuracy and good agreement, with an overall accuracy and kappa values of 0.72, 0.98, 0.711, and 0.955 for the quadruple classification, dichotomous classification, quadruple classification Kappa value, and dichotomous classification Kappa value, respectively. Table 4 provide the accuracy and consistency of ResNet101 evaluation.

<table>
<thead>
<tr>
<th>Classification accuracy</th>
<th>Classification accuracy</th>
<th>Kappa value</th>
<th>Kappa value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(quadruple classification)</td>
<td>(dichotomous classification)</td>
<td>(quadruple classification)</td>
<td>(dichotomous classification)</td>
</tr>
<tr>
<td>Model vs. Clinical</td>
<td>0.72</td>
<td>0.98</td>
<td>0.711</td>
</tr>
</tbody>
</table>

**Visual heat map analysis**
The aforementioned network could present visualization results, and generate the visualization result map of the ResNet101 model classification output, that is, the category activation visualization heat map. Different stages of pneumoconiosis and their corresponding category activation visualizations are presented in Supplemental Fig. 5-a Fig. 5-d. We evaluated the learning ability of the model using a heat map and the overlay presentation of original images. For relevant features of pneumoconiosis images, brighter the color, higher the possible value of the predicted pneumoconiosis lesion. The red area represented the closest value to which the network predicted relevant features of the pneumoconiosis image. This corresponds to the staging of pneumoconiosis given by our radiologists.

**DISCUSSION**

This study explored the feasibility of deep learning-based techniques for automatically assessing pneumoconiosis in DR images. It not only screened participants for pneumoconiosis but also performed a quadruple classification of the disease. We tested and compared three CNN models to select the final CNN architecture, including ResNet50, ResNet101, DenseNet, and ResNet101 as the optimum networks. In addition, three board-certified radiologists independently interpreted the images in the test dataset, and their performances were compared with that of the computerized protocol.

According to ILO standards, pneumoconiosis diagnosis and staging is based on the filling degree of small opacities, appearance of large opacities, and aggregation of small opacities on DR. This conventional diagnostic process is subjective and time-consuming, thus leading to misclassification and unreliable results. The aforementioned problem worsens and becomes frequent while conducting pneumoconiosis screening programs in underdeveloped areas. We developed a deep learning approach to automatically evaluate pneumoconiosis on chest DR, which provided screening results and the stages of pneumoconiosis. Our approach provided a detailed visual interpretation of the predictions, One can have more confidence in the ability of the classifier in a black box problem, through such visualization methods. [29]. The less tolerance of clinical diagnoses to incorrect predictions necessitates interpretation to help physicians make decisions. The primary difference between our proposed deep learning-based pneumoconiosis screening model and other models is that the deep learning technique can directly extract features from the training data, thus significantly reducing the training workload of feature extraction and the impact of human intervention. In addition, we not only screened participants for pneumoconiosis but also performed the four classifications to guide the diagnosis and treatment. Furthermore, this study increased the transparency of pneumoconiosis classification by providing a visual interpretation of the model to improve its reliability. Our findings may also serve as a roadmap for unlocking the “black box” principle of artificial intelligence for image analysis tasks in other medical fields. Eventually, through a comparative analysis of the model and clinical assessment, this study further confirmed the clinical utility of the model in pneumoconiosis classification. We used three convolutional neural models to screen and stage pneumoconiosis, which was divided into four categories ranging from 0 to III. ResNet101 was the most suitable model for pneumoconiosis screening and classification, with AUC values of 1.0, 0.90, 0.89, and 0.94 for category 0, category 1, category 2, category 3, respectively. With AUC values of 0.93 and 0.94, the overall model displayed better effectiveness metrics than Resnet50 and DenseNet, respectively. ResNet101 outperformed models used in previous studies. Okumura et al. [30] developed a CAD system based on the rule of ILO and an artificial neural network (ANN) power spectrum analysis of pneumoconiosis. The results of classifying normal pneumoconiosis with abnormal pneumoconiosis demonstrated mean AUCs of 0.93 and 0.72 on chest X-ray films in the highest category (severe pneumoconiosis) and lowest category (early pneumoconiosis), respectively. Subsequently, Okumura et al. [31] developed a CAD system based on ANN classification of the textural features of pneumoconiosis chest films. The image database consisted of 36 chest films divided into four categories, ranging from 0 to 3. The AUC values for category 3 pneumoconiosis and category 0 pneumoconiosis were 0.89 ± 0.09 and 0.84 ± 0.12, correspondingly. We compared the accuracy and consistency of deep learning models with those of the relative gold standard, where that the model classification accuracy was 0.80, compared with the clinical standard classification. The kappa value and accuracy of the quadruple classification were 0.733. The accuracy and kappa value of the two classifications were 0.98 and 0.931, respectively. Wang X et al. [32] explored the potential of deep learning in assessing
pneumoconiosis, as revealed by digital chest radiographs, and compared their performance with radiologists. They used the inception-V3 Network. The model's AUC was 0.878, whereas that of the two radiologists were 0.668 and 0.772, respectively. Moreover, the readers displayed moderate agreement ($\kappa = 0.423, p < 0.001$). However, Wang et al. only screened patients for pneumoconiosis. Only few researchers have performed a quadruple classification of pneumoconiosis [33]. We performed screening for pneumoconiosis and staging tasks, where our proposed model performed better.

Our study had some limitations. First, the dataset only included chest X-ray images. Despite chest radiography being the standard method for pneumoconiosis diagnosis, X-ray chest films have other characteristics, such as insufficient resolution, overlap effect, and factors with skeletal (noise) images. The next step involved designing a network structure and data preprocessing process that was suitable for the aforementioned characteristics. CT usually provides details of shadowed areas within the lungs. Some countries have introduced regulations for high-resolution CT as a diagnostic criterion for pneumoconiosis. Considering this future trend, shadowed areas should be tested in future studies as patients with pneumoconiosis cannot undergo pathology tests and open-chest examinations. The criteria used in this study were relatively subjective gold standards. We selected patients with pneumoconiosis and healthy individuals for model training. The diagnosis of patients with pneumoconiosis should be combined with relevant dust reception history and laboratory tests to provide a definite diagnosis. However, it is not possible to distinguish pneumoconiosis from other lung diseases with similar imaging signs. This will be the next research direction to open new horizons.

In conclusion, this study proposed a deep learning-based model for detecting and classifying pneumoconiosis cases using DR images. We proposed a fully automated end-to-end architecture model that did not require manual feature extraction. It could perform binary and multi-classification tasks with accuracies of 98% and 72%, respectively. This model could further simulate the diagnostic behavior of the radiologists. The system can be used in remote areas of pneumoconiosis-affected countries to overcome the shortage of radiologists. In addition, these models can be used to diagnose other diseases associated with the chest, including tuberculosis and pneumonia. Therefore, it is worthwhile to develop further deep learning solutions for pneumoconiosis screening and classification in clinical practice.

**Abbreviations**

- ILO International Labor Organization
- CAD computer-aided diagnosis
- CNNs convolutional neural networks
- DR digital radiography
- ANN artificial neural network

**Declarations**

**Acknowledgements**

Not applicable

**Authors' contributions**

Bowen Zheng: Methodology, Software, Writing - Original Draft; Long Li: Conceptualization, Methodology; Fengxia Zeng: Formal analysis; Tianqiong Wu: Data Curation, Investigation; Xiaoke Cheng: Data Curation, Investigation; Yuli Peng: Data Curation, Investigation; Yonliang Zhang: Data Curation, Investigation; Yuanlin Xie: Resources; Wei Yi: Resources; Weiguo Chen: Conceptualization; Genggeng Qin: Conceptualization, Data Curation; Jiefang Wu: Conceptualization
Funding

Guangzhou Health and Health Science and Technology Project (20231A011065),

National Natural Science Foundation of China (82171929), National Key R&D Program of China (2019YFC0117301),
National Key R&D Program of China (2019YFC0121903), Natural Science Foundation of Guangdong Province
(2018A0303130215)

Availability of data and materials

The dataset analysed for the manuscript is available upon reasonable request. The datasets analyzed in this study are
available from the corresponding author on request.

Ethics statement

This study was approved by the Medical Neighborhood Committee of 12th people's Hospital of Guangzhou . All participants
gave informed consent. All methods were performed in accordance with the relevant guidelines and regulations.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

1Department of Radiology, 12th people's Hospital of Guangzhou , Guangzhou 510620, China

2Department of Radiology, Nan fang Hospital, Southern Medical University, Guangzhou 510515, China

3Department of Radiology, San shui District Institute for Disease Control and Prevention Foshan Guangdong 528100
China

4Department of Radiology, The Third People's Hospital of Yunnan Province Yunnan 650010 China

References

1. Blackley, D. J., Halldin, C. N. & Laney, A. S. Continued increase in prevalence of coal workers' pneumoconiosis in the

28727677; PMCID: PMC5657940.

and new exposures to respirable crystalline silica — United States, 2001-2010. MMWR Morb Mortal Wkly Rep


Figures
Schematic flow chart illustrating the procedure for the four classifications of pneumoconiosis.

**Figure 1**

Category activation visualization heat map
Figure 2

shows: The accuracy and loss rate of the training process, the Abscissa represents the smooth times, and the ordinate distinguishes the accuracy and the loss rate respectively, which increases over time, the accuracy on the training set and the verification set will increase, and the loss rate will decrease. As expected, there is a reduction of loss over the course of training as accuracy improves. The loss on the validation is similar to the training, which indicates that there is no appreciable overfitting. These training curves are used for model selection. In this case, the best performing model at epoch 1000 was used on the test data for final assessment. Val = validation.
Figure 3

a Evaluation index of ResNet50 convolution neural network

b Evaluation index of ResNet101 convolution neural network

c Evaluation index of DenseNet convolution neural network
Figure 4

Receiver operating characteristic (ROC) curve of different CNNs. A ROC curve of ResNet50. B ROC curve of ResNet101. C ROC curve of DenseNet
Figure 5

a X-ray images and the corresponding heat maps: (a) pneumoconiosis stage 0 of X-ray image, (b) heat map of (a).

b X-ray images and the corresponding heat maps: (a) pneumoconiosis stage I of X-ray image, (b) heat map of (a).

c X-ray images and the corresponding heat maps: (a) pneumoconiosis stage II of X-ray image, (b) heat map of (a).

d X-ray images and the corresponding heat maps: (a) pneumoconiosis stage III of X-ray image, (b) heat map of (a).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryfile.docx