Development of Automated Diagnostic Tools for Pneumoconiosis Detection from Chest X-Ray Radiographs

The final report prepared for Coal Services Health and Safety Trust

Yulia Arzhaeva, Dadong Wang, Liton Devnath, Saeed Amirgholipour, Rhiannon McBean, James Hillhouse, Suhuai Luo, David Meredith, Katrina Newbigin and Deborah Yates

Coal Services Health and Safety Trust Project No. 20647
CSIRO Report No. EP192938
15 May 2019

Health & Safety Trust

ST VINCENT'S HOSPITAL SYDNEY

wesley medical imaging
advanced diagnostics, better care

Quantitative Imaging Research Team, Data61, CSIRO

Contents

Figures

Tables

Acknowledgments

The authors would like to thank Coal Services Health for providing de-identified sample X-ray images from their image database, and radiologist Dr Katrina Newbigin from Wesley Medical Imaging for providing examples of nodular dust related lung disease including cases of coal worker's pneumoconiosis and silicosis from their database at the Wesley Hospital Brisbane. Thanks are also due to Dr James Hillhouse, Brian Sorensen, Lorretta Jacob, Hiep Pham and Andy Young of St Vincent's Hospital for providing X-Ray images for this study. The authors would also like to extend their gratitude and appreciation to Prof. Robert Cohen, Director of Occupational Lung Disease from Northwestern University in the US, for providing advice at the initial stage of the project.

Executive Summary

Background and Objectives of the Project

Pneumoconiosis is caused by long-term inhalation of respirable dust, such as coal, asbestos, and silica, and that from the inhalation of coal dust is more commonly known as black lung. It is characterized by declining lung function and has no cure. In Queensland, Australia, about 105 cases of mine dust lung diseases have been diagnosed since 1984 [1]. In 2017, the first case was identified in an NSW coal mine and broke the non-occurrence record that the state had been justifiably proud of since 1970s [2]. One case is too many. It is reported that pneumoconiosis kills about 6,000 coal workers in China each year [3]; and in the US, more than 10% of examined underground coal miners with 25 or more years of experience were diagnosed with pneumoconiosis [4]. Pneumoconiosis caused 69,377 deaths during 1970-2004 [5], and about 21,600 people died of pneumoconiosis globally in 2017 alone [6]. A report shows that poor dust control is to blame for the re-emergence of pneumoconiosis in Queensland, and patchy medical screening has failed in the early detection of this potentially fatal disease [7]. A 2018 study by the National Institute of Occupational Safety and Health (NIOSH) also shows a resurgence of this disease in the US [4].

For pneumoconiosis screening, chest X-ray images (radiographs) are acceptable, widely available and relatively inexpensive. The main manifestation of pneumoconiosis is the presence of small, regular and irregular, opacities in the lung parenchyma. In the International Labour Organization (ILO) Classification System, small opacities are categorised according to their shape (round or irregular) and size (ranging from <1.5 to >10 mm). The ILO Classification System requires a B-reader to compare chest X-rays to reference standards to provide a score on a 12-point grading scale [10].

There is no national approach to health screening of coal miners in Australia. In NSW, a chest X-ray is normally recommended every 6 years for mine site workers but is not mandatory. X-rays are read by radiologists who are familiar with the ILO classification but may not be certified B-readers. The current practice in Queensland is considered the most stringent of health screening programs across the country, after undergoing significant reform since 2016. In Queensland, coal miners considered at risk of dust exposure are required to undergo pre-employment chest X-rays, followed by routine X-ray screenings after the employment, and each X-ray requires two B-readers to review. However, the insensitivity of chest radiographs for the detection of early or moderate pneumoconiosis limits their efficacy in screening [8]. This also leads to low sensitivity and specificity of chest X-rays when read by a radiologist who is qualified as a B-reader, especially for the detection of pneumoconiosis at an early stage of the disease. Inter- and intra-reader variability in chest radiography has been acknowledged ever since chest radiography was first used to identify and classify pneumoconiosis. Another limiting factor in screening programs is that there are only 71 doctors from outside the US who are currently certified to the B-reader standard to identify pneumoconiosis in chest radiographs [9]. This indicates that the B-readers are in very short supply, and in some cases, a large backlog of X-rays could occur. Additionally, the false positive rate for radiologists reading X-rays has been reported as between 23-27% [11]. To date, there has been a lack of systematic, automated, and objective systems for detecting the presence

and assessing the progression of pneumoconiosis for individual coal miners other than by expert radiologists.

Past methods for automated detection of pneumoconiosis include using classical image analysis to extract a set of handcrafted features from each lung field and zone. The features were extracted based on pixel intensities, c-occurrence matrix and frequency domain. A subset of these features was selected as input to train Support Vector Machine (SVM) classifiers to predict whether or not a region of interest in an X-ray contained any abnormalities [12]. This required substantial work to extract the handcrafted features and employ various methods to select discriminative features to build the SVM. In the last five years, due to advances in deep learning, there have been many successful applications of deep learning to image classification and abnormality detection problems. Some deep learning algorithms even "go beyond" the performance of medical professionals in a variety of medical imaging tasks. For example, CheXNet [13] was developed by Stanford Machine Learning Group to detect pneumonia from chest X-rays. The core of CheXNet is a 121-layer dense convolutional neural network (DenseNet) [14] that uses a chest X-ray image as input and generates the probability of pneumonia along with a heat map localizing the areas of the image most indicative of pneumonia. The CheXNet was trained on ChestX-ray14 image database [15] with over 100,000 X-ray images of 14 different thoracic diseases acquired from more than 30,000 unique patients. When training CheXNet, all pneumonia X-ray images from this database were labelled as positives and the rest of the images were deemed as negatives. Apart from the large training dataset, 420 chest X-rays were used for testing. The testing results showed that the CheXNet outperformed the average radiologist on pneumonia detection.

With pneumoconiosis, the low incidence of this disease and restrictions on sharing patient data means that the number of available chest X-rays may not be sufficient for developing a deep learning model for automated detection of the disease. Therefore, detecting pneumoconiosis in chest X-rays remains a challenging task that relies on the availability of expert radiologists.

In collaboration with Coal Services Health (CSH), Wesley Medical Imaging at Queensland, and St Vincent's Hospital at Sydney, this project aimed at addressing the above problems by developing Computer-Aided Diagnosis (CAD) tools for automated pneumoconiosis detection using chest X-rays.

Achievements

We have evaluated different approaches including statistical image analysis, classical machine learning methods, and some state-of-the-art deep learning models. We have also developed a customised cascade learning model for the automated detection of pneumoconiosis using both real and synthetic pneumoconiosis radiographs. With the cascade learning, we employed (1) a machine learning based pixel classifier with post processing for lung field segmentation, (2) Cycle-Consistent Adversarial Networks (CycleGAN) [16] for generating abundant lung field images for training, and (3) an image classier using a 15-layer Convolutional Neural Network (CNN) trained with the CycleGAN generated and real chest X-ray images. These machine learning models are trained sequentially one at a time, then join forces to form a cascade machine learning workflow. Experiments are conducted to compare the classification results from several state-of-the-art machine learning models and ours. Our proposed model outperforms the others and achieves an overall classification accuracy of 90.24%, a sensitivity of 93.33%, and a specificity of 88.46% for detecting pneumoconiosis. The experiments also show improved performance on the pneumoconiosis detection by leveraging the synthetic images and demonstrate that the cascade

learning model can be potentially used as a tool for the pre-screening of pneumoconiosis. We have also developed a web-based demo to show how our proposed machine learning model works.

Below is a summary list of what we have achieved during this project:

- Obtained Human Research Ethics Approvals from CSIRO and St Vincent's Hospital.
- Collected chest X-ray images from CSH, ILO, Wesley Medical Imaging, St Vincent's Hospital, and the USA National Institute of Health.
- Developed and implemented algorithms for lung field and zone segmentation for both digital and digitised analogue X-ray images.
- Developed statistical image analysis methods and evaluated some classical machine learning algorithms for automated detection of pneumoconiosis:
    - Using Multi-Layer Perceptron (MLP) based method for automated detection of pneumoconiosis which showed 71.11% pneumoconiosis detection accuracy.
    - Developed statistical image analysis methods for pneumoconiosis detection, including the automated detection and quantification of opacities for each zone. The quantitative analysis results can then be used for the classification of ILO categories of pneumoconiosis. Using custom image features based on evaluating disk-like structures within the lung fields, an accuracy of 76.9% for the automated detection of pneumoconiosis was achieved with a Ridge classifier.
    - Developed a three-class classification algorithm that uses texture image features and a neural network classifier to differentiate among normal, early stage pneumoconiosis and severe pneumoconiosis cases. The algorithm achieved a reasonably good recall for normal and severe pneumoconiosis images – 86% and 90%, respectively, but a weak recall of 38% for early stage pneumoconiosis images.
    - The binary classification of the same algorithm – normal images vs. pneumoconiosis images – achieved 83% accuracy, 85% sensitivity, and 82% specificity.
- Investigated machine learning based one-class classification which has been used to deal with class imbalance - significantly more training images from one class are available than those from other classes. In this study, the number of normal X-ray images are much more than that of pneumoconiosis X-rays. We have conducted the following experiments:
    - Using One-Class Support Vector Machines (OC-SVM) with raw chest X-ray images as input, the best sensitivity, specificity and accuracy for discriminating between normal and pneumoconiosis X-rays are 73.3%, 92.31% and 73.17%, respectively.
    - Using Isolation Forest with raw chest X-ray images as input, the best sensitivity, specificity and accuracy achieved are 93.33%, 88.46% and 68.29%, respectively.
    - Using the hybrid model of Autoencoder [17] and a Feed Forward Neural Network classifier, the best sensitivity, specificity and accuracy produced are 60%, 88.46% and 68.29%, respectively.

- Investigated machine learning based two-class classification to identify normal and pneumoconiosis X-rays by training classifiers with equal number of training images from each class. We carried out the following experiments:
    - Using Support Vector Machines (SVM) with raw chest X-rays as input, the best sensitivity, specificity and accuracy we have obtained are 33.33%, 88.46% and 68.29%, respectively.
    - Using the hybrid model of Autoencoder and SVM, the best sensitivity, specificity and overall accuracy we have produced are 93.33%, 76.92% and 73.17%, respectively.
    - Using the hybrid model of CheXNet and SVM with deep features as input, the best sensitivity, specificity and accuracy we have obtained are 80%, 80.77% and 78.05%, respectively.
    - Using the pre-trained DenseNet with local image patches as input and aggregating the local classification results into a single image label, we have obtained a sensitivity of 93.33%, specificity of 80.77%, and overall accuracy of 85.37%.
- Developed a cascade deep learning model which outperforms others and achieved overall classification accuracy of 90.24%, a specificity of 88.46% and a sensitivity of 93.33% for detecting pneumoconiosis.
- Implemented a web application "Black Lung Prediction Tool" that uses the cascade learning model to classify 41 chest radiographs available on the web site. The web demo can be found at http://confederate.csiro.au/.

Limitations and Suggestions

Due to the low incidence of pneumoconiosis in Australia we were able to validate our tool only with a limited number of chest X-rays with pneumoconiosis. To make our tool more robust and suitable for clinical use, we suggest:

- To work with radiologists on additional acquisition of chest X-rays with features of pneumoconiosis;
- To improve our automated system for detecting and grading pneumoconiosis into different categories of severity when more chest radiographs become available; and
- To set up a pilot study where our tool is being used in a clinical setting in parallel with B-readers. This can be done in collaboration with Coal Services Health to examine their chest radiographs currently reread by an independent radiologist every quarter. This may take form of a web-based (secure) application or a stand-alone tool installed on Coal Services Health hardware. This pilot software can be used retrospectively on the chest radiographs. The feedback from the pilot study can be used to further improve functionality and performance of our automated pneumoconiosis detection tool.

1 Human Research Ethics Approval

We prepared the study protocol for this project [18] and submitted it for ethical and scientific review by St Vincent's Hospital Research Office. It was granted ethical and scientific approval for this multi-centre project on 15/08/2017. Also, this project received an approval from CSIRO Health and Medical Human Research Ethics Committee on 12/10/2016 [19]. All research panels deemed it a low/negligible risk project.

2 Image Data Collection

We have collaborated with various organisations to obtain image datasets and associated labels to be used in the development of the automated diagnostic system. We have also used publicly available NIOSH teaching chest X-ray datasets to develop parts of the system. All radiographs used in this study are posterior-anterior (PA) radiographs, some of which are fully digital, while some are digitized films.

2.1 Chest X-Ray Images from International Labour Organisation (ILO)

We have purchased a digital set of 22 ILO Standard Radiographs in DICOM format. This set is used in the ILO Classification System for Pneumoconiosis on Chest Radiographs (ILO Classification). ILO Classification protocol recommends classifying a subject's radiograph by visually comparing it with ILO Standard Radiographs. For the purpose of this project, we have selected 17 chest radiographs out of the 22. The selected images depict complete lung fields – either normal, or with small parenchymal abnormalities consistent with pneumoconiosis. In addition to this dataset, we have downloaded the online B Reader Syllabus by the National Institute for Occupational Safety and Health intended for preparing doctors to take the ILO Classification exam (NIOSH (2000)) [20]. 41 teaching images from this resource were selected for our study, using the same criteria as for selecting ILO Standard Radiographs. These data are used in this project for training the automated pneumoconiosis diagnostic system.

2.2 Chest X-Ray Images from Wesley Medical Imaging, QLD

We have collected 62 chest X-rays belonging to normal individuals (56 males and 6 females, all de-identified), and 36 chest X-rays with small parenchymal opacities consistent with pneumoconiosis which belong to 27 de-identified male individuals. These data are also used in this study for training and evaluation of the automated pneumoconiosis diagnostic system.

2.3 Chest X-Ray Images from National Institute of Health, the USA

We have also downloaded the ChestX-ray14 dataset from National Institute of Health that was made publicly available in September 2017. This data set includes over 100,000 X-ray images of more than 30,000 unique patients collected from a hospital Picture Archiving and Communication System (PACS) with automatically text-mined image labels from their associated radiological reports [15]. These data are used in this study for exploring deep learning-based methods for the automated detection of pneumoconiosis chest X-rays.

2.4 X-Ray Images from CSH

We have acquired 511 chest X-ray images from Coal Services Health (CSH), including 505 chest X-rays exhibiting no signs of pneumoconiosis, 5 chest X-rays classified as ILO 0/1 that might have features consistent with pneumoconiosis, and one X-ray classified as ILO 2/2.

2.5 X-Ray Images from JSRT Database

JSRT database [21] is a public database that has been previously used in many lung segmentation studies [24]. We downloaded this database to train the lung segmentation algorithm described in Section 3. JSRT contains 247 digitized chest X-rays with annotated lung masks ("gold standards").

3 Lung Field Segmentation

Delineation of the lung fields in chest X-rays, otherwise called lung segmentation, is a pre-requisite for the most of computer-aided evaluation systems for chest X-rays. We have implemented a previously published algorithm that employs Pixel Classification to distinguish between lung and non-lung areas in a radiograph. Further we have introduced necessary modifications to the algorithm to improve its performance on fully digital radiographs.

3.1 Pixel Classification

3.1.1 Algorithm description and implementation

Pixel Classification with post processing (PC) was first described and compared with the previous state-of-the-art segmentation techniques in [23]. In 2015 it was still compared favourably with other lung segmentation algorithms validated on the same JSRT database [24]. PC yielded around 95% overlap score with the JSRT gold standard lung masks.

PC is a pattern recognition technique, where training and testing stages can be distinguished. In the training stage an image is resized to a working resolution and subsampled. For each sample in a subsampled image a set of features are extracted. The features are computed from a neighbourhood centred on this sample and are devised to characterise local image structures. It is assumed that small neighbourhoods from inside the lung fields have a distinctively different appearance to small neighbourhoods outside the lungs. In this algorithm, the output of Gaussian derivative filters at multiple scales are used to characterize local image structures. In addition, X and Y coordinates of each sample are included in the feature set. Each such feature set has a corresponding label, 0 – if a pixel belongs to image background, 1 – for a pixel in the right lung, and 2 – for a pixel in the left lung. Next, a K-Nearest Neighbour (k-NN) classifier is trained with the feature sets and the corresponding labels, learning how to map pixel features to class labels. In the end of the training stage a classifier can compute a probability that a new input pixel belongs to a certain object class (image background, right lung or left lung).

In the testing stage, a new unknown image is resized to the working resolution, then, the same feature set is computed for each pixel in the image. A trained k-NN classifier takes each pixel's feature set as an input and computes a probability for that pixel to belong to each of the three classes, p_0, p_1 and p_2. This allows us to create a lung probability map P. It has the same size as the test image, and its pixel values, $p(x,y) = p_1(x,y) + p_2(x,y)$, define a probability that a pixel belongs to one of the lung fields (Figure 1(b)). The obvious way to turn the probability map into a lung mask is thresholding it at a probability of 0.5, meaning that every pixel that received a probability greater than 0.5 is assumed to be a lung pixel. However, in this way, lung segmentation will always contain clouds of isolated pixels near the lungs' border. To ensure connectedness of the lung fields, we blur the probability map first, then perform thresholding at 0.5. The two largest connected objects in the resulting binary mask are labelled as 1 (the right lung) and 2 (the left lung), and holes in the masks are filled (Figure 1(c)).

(a) (b) (c)

Figure 1 A chest radiograph (a) with a corresponding probability map (b) computed using Pixel Classification, and a labelled lung mask (c) obtained by post-processing the probability map.

This algorithm was implemented in Python using image analysis and machine learning toolkits **SimpleITK** [25] and **scikit-learn** [26].

3.1.2 Training and validation

The JSRT database with its "gold standard" lung masks was used for training PC and validating its performance. 124 randomly selected radiographs from JSRT database were used to train the algorithm, and the rest of the database was used to test it. We obtained very similar results to those which were published in [23] – an average 95% overlap score with 1% standard deviation on 123 test X-rays.

3.2 Application to Digital X-rays

Next, we applied the PC algorithm trained with the full JSRT database of 247 images to radiographs from the Wesley Medical Imaging dataset (Section 2.2) as well as to the radiographs from the ChestX-Ray14 dataset (Section 2.3). The resulting segmentation was not satisfactory, as in many cases the algorithm failed to correctly separate the lung fields from other structures in the images. The main reason for this, in our opinion, was different technologies employed to acquire JSRT radiographs and the radiographs in the other two datasets. JSRT radiographs were digitized copies of film x-rays while the other radiographs were fully digital images acquired using digital radiography units. Such images look differently and have different pixel intensity distributions. To illustrate this, Figure 2(a) shows a digitized radiograph from JSRT database and the histogram of its intensity values, and Figure 2(b) shows a digital radiograph from the Wesley Medical Imaging dataset and its corresponding histogram.

One way to successfully apply the PC lung segmentation algorithm to digital data is to train it on images similar to the test data, i.e. on some other set of digital radiographs. For this, we would need to find or manually produced "gold standard" lung masks for training. Alternatively, the test data could be made more similar to the already available training data by means of image processing. Since obtaining lung masks for training was not a feasible option for us, we opted to simulate a digitized "look" on the images from our digital datasets.

We have employed histogram matching to normalize the pixel intensities of digital images based on the pixel intensity values of digitized radiographs. It is based on the work published in [27]. We have randomly selected ten reference images from JSRT database. A digital radiograph was then matched to each of the reference images, and an average of the ten resulting images was computed. The intensity histogram of such a simulated image resembled more closely the histograms of digitized radiographs. To illustrate this, an image in Figure 2(c) shows the result of histogram matching between a digital radiograph in Figure 2(b) and the ten reference JSRT images. Figures 2(e) and 2(f) show the corresponding intensity histograms. Note that the histogram in Figure 2(f) is noticeably more similar to the histogram of a digitized chest X-ray (Figures 2(a) and 2(d)).

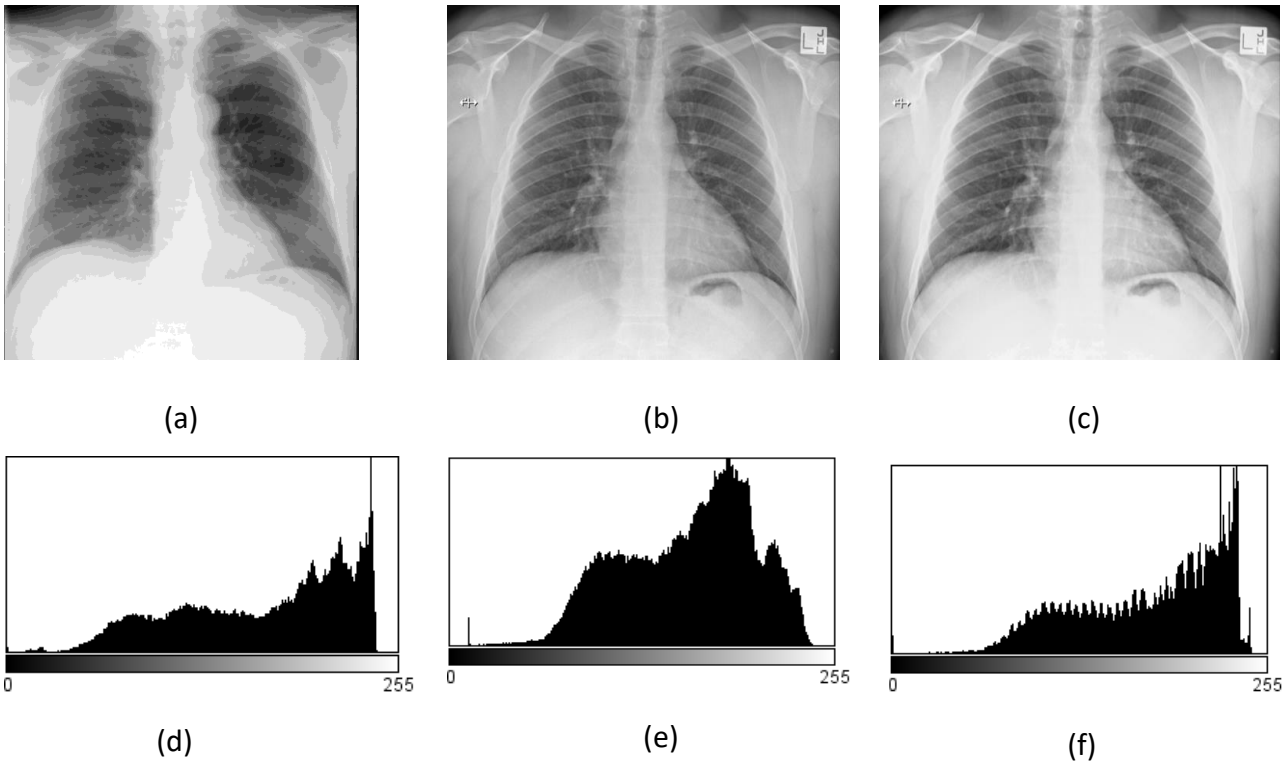


Figure 2 A digitized X-ray from JSRT database (a) and its intensity histogram (d); a digital X-ray from Wesley Medical Imaging (b) and its intensity histogram (e); and (c) the modified version of the X-ray in (b), with its intensity histogram (f) that now resembles the histogram in (d).

After applying histogram matching to all the radiographs from the Wesley Medical Imaging dataset and the ChestX-Ray14 database, it became possible to compute lung masks successfully by using Pixel Classification trained on JSRT database. Visual inspection suggested that the computer-generated masks have minimal errors and cover the lung fields sufficiently for our goal to automatically detect signs of pneumoconiosis inside the lungs. Figure 3 demonstrates the four examples of the lung fields segmented from the Wesley Medical Imaging dataset radiographs using the PC lung segmentation with histogram matching.

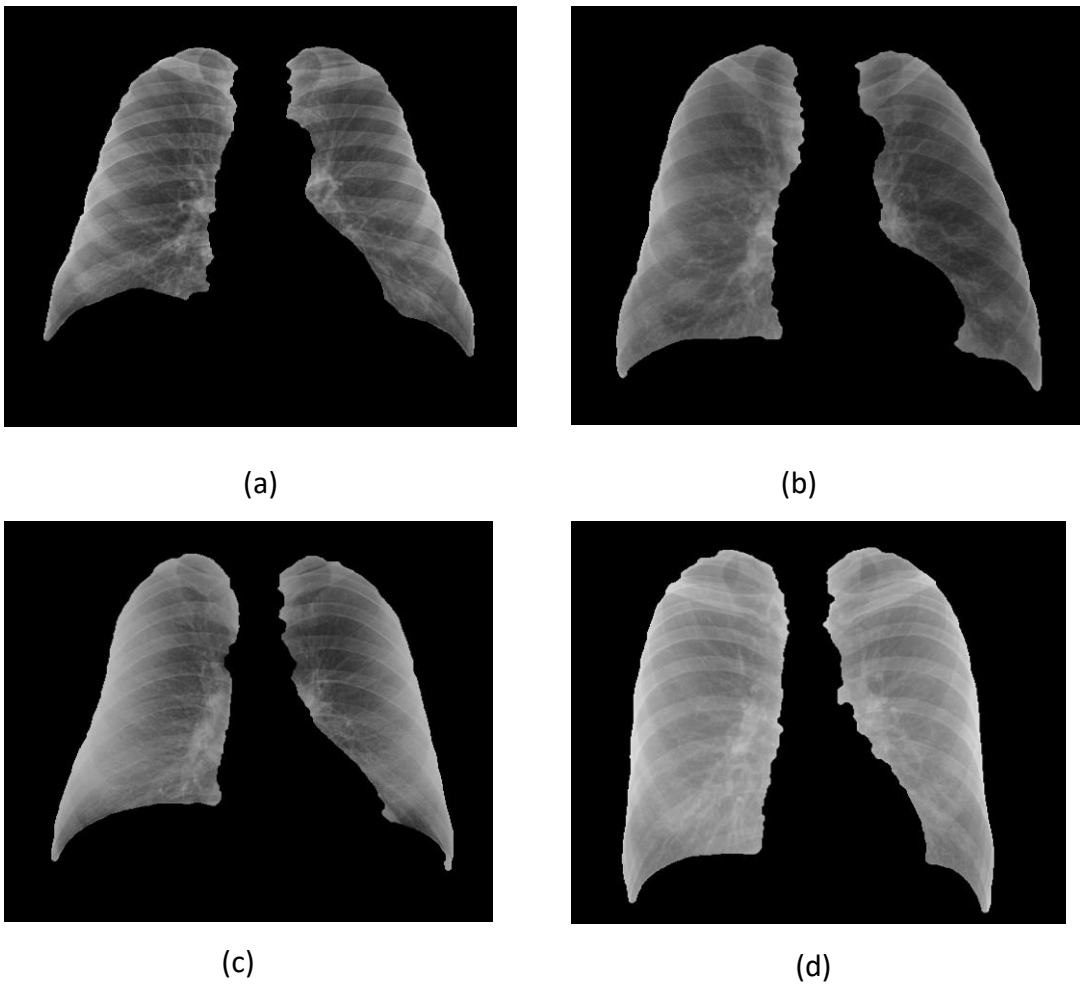


Figure 3 Examples of the lung fields automatically segmented from the digital radiographs obtained from Wesley Medical Imaging. Images (a) and (b) depict normal lungs, while (c) and (d) depict lungs with some features of pneumoconiosis.

3.3 Lung Zones

According to the ILO Classification system [10], the profusion of small opacities is determined over each of the six lung zones. Therefore, the last step in our lung segmentation algorithm performs division of the lung fields into zones. The automated zone division replicates the ILO Classification System [10] that states that "each field is divided into three zones by horizontal lines drawn at approximately one-third and two-thirds of the vertical distance between the lung apices and the domes of the diaphragm". The algorithm assigns each zone a label, from 1 to 6, for an easy reference. Figure 4 demonstrates a few examples of the automated zone division.

Figure 4 Examples of the lung fields automatically divided into six zones according to ILO Classification System. Zones' labels, from 1 to 6, are mapped to colours for the reader's convenience.

4 Statistical Image Analysis and Classical Machine Learning Based Automated detection of Pneumoconiosis

In this section, we report the experimental results obtained by employing hand crafted features and classical machine learning approaches like Support Vector Machine (SVM), Multi-Layer Perceptron (MLP), Perceptron, K-Nearest Neighbours algorithm (K-NN), Ridge Classifier, Random Forest. The handcrafted features are extracted from X-Ray images using statistical image analysis algorithms. We have evaluated various classical machine learning schemes for the detection of pneumoconiosis. One is based on one-class classification and the other is based on two-class classification.

4.1 One-Class Classification Based on Classical Machine Learning

With the images provided by Coal Services Health, only one was clearly an X-ray with pneumoconiosis (ILO 2/2). To leverage the availability of large number of normal images and small number of pneumoconiosis images, one-class classification is a straightforward choice. With this method, a classifier is trained to learn and summarise the features of normal images in a training dataset which contains normal images only. After training the classifier, we expect that the pneumoconiosis images can be identified as anomalies or outliers by the classifier, in comparison with the normal X-ray images used for the training.

Based on the above idea, we tried three combinations of feature extraction methods and classifiers. The experimental settings and results are reported below.

4.1.1 Support Vector Machines

Support Vector Machines (SVMs) are a set of supervised learning methods used for classification and regression. It was first proposed by Vapnik et al. for classification [28] and has become an intensive research area in the last few decades. The basic idea of SVM for solving classification problems is to construct an optimal hyperplane or a set of optimal hyperplanes in a high dimensional feature space. Because of the nature of the feature space in which these hyperplanes are generated, SVM can exhibit a large degree of flexibility in handling classification tasks of varied complexities. They have been used widely in various applications such as face recognition [29]. When performing classification tasks, SVM is a discriminative classifier defined by one, or a set of, optimal hyperplanes which are constructed during training. Therefore, the SVM algorithms are designed to find the hyperplane that gives the largest minimum margin to the training samples. The two key elements in SVM are (1) a general-purpose learning machine, and (2) a problem specific function, which can be a linear, polynomial, sigmoid or Radial Basis Function (RBF). The flexibility of these kernel functions enables the SVM to explore a wide variety of hypothesis spaces. SVM can be formulated with two things: the hypothesis spaces and the loss functions.

SVM is effective in high dimensional spaces, even in cases where number of dimensions is greater than the number of samples.

There are two types of SVM for classification, namely C-SVM [30] and nu-SVM [31]. Their trainings involve different error functions to minimise. 'C' and 'nu' are both regularisation parameters used to improve the accuracy of the classification by helping implement a penalty on the misclassifications that are performed while separating the classes. 'C' ranges from 0 to infinity, the larger the C, the more the error is penalized. 'nu' is between 0 and 1, it is related to the ratio of support vectors and the ratio of the training error and serves as an upper bound on the fraction of margin errors and a lower bound on the fraction of support vectors.

4.1.2 Autoencoder

In our study, we have also used Autoencoder for feature extraction. It is a specific type of feedforward neural networks where the input is the same as the output. They compress the input into a lower-dimensional code and then reconstruct the output from the code. The code is a compact "summary" or "compression" of the input, also called the latent-space representation.

An Autoencoder consists of 3 components: encoder, code, and decoder as shown in the following figure. The encoder compresses the input and produces the code, and the decoder then reconstructs the input only using this code. To build an Autoencoder, we need an encoding method, a decoding method, and a loss function to compare the output with the target which is the same as the input.

The Autoencoder is mainly a dimensionality reduction or compression algorithm with the following properties:

a) Data-specific: An Autoencoder is only able to meaningfully compress data similar to what they have been trained on. Since it learns features specific to a given training dataset, we cannot expect an Autoencoder trained on one type of images to compress a different type of images.
b) Lossy: The output of an Autoencoder will not be the same as the input, it will be a close but degraded representation.
c) Unsupervised: To train an Autoencoder, we just use the raw input data as its input and output. The Autoencoder is considered as an unsupervised learning technique since it does not require explicit labels to train on. They are self-supervised because they generate their labels from the training data.

As a popular neural network model that learns hidden representations of unlabelled data, an Autoencoder [32, 33] and its variants [34, 35] can also be used as a feature extractor to learn a representation of image data. In this study, we use Autoencoder to learn image features and then feed the features into Neural Networks (NN) for the detection of pneumoconiosis X-rays.

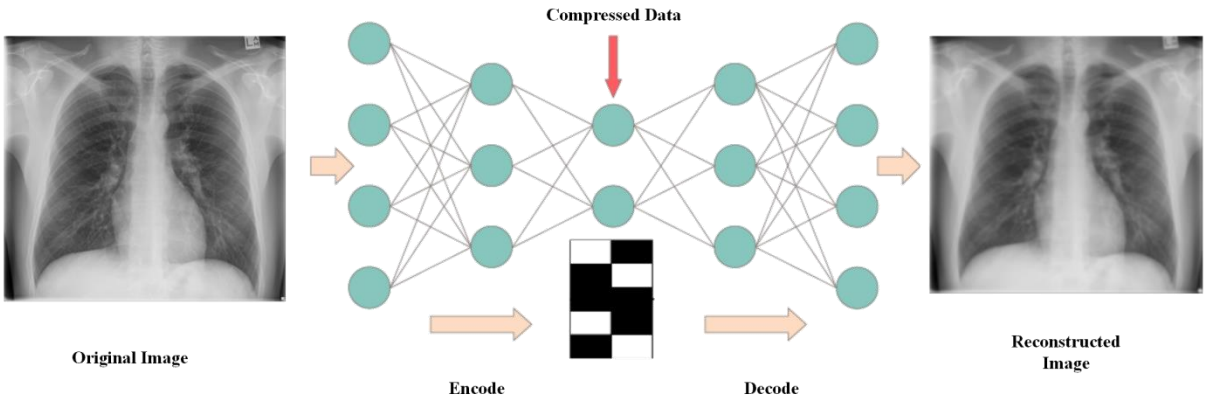


Figure 5 The Autoencoder network architecture - an input Lung image is fed to the encoder and the network is trained to encode and decode the image

4.1.3 Experiments using OC-SVM with Raw Chest X-ray Images as Input

In these experiments, the masked chest X-rays are used as the input of an OC-SVM. Using the lung field segmentation algorithm from Section 3, we successfully generated the lung masks of normal X-ray images. The masked radiographs are generated by multiplying the raw chest-X-ray images and their lung masks. The lung masks are automatically retrieved using a pixel-based classification method as described in Section 3. The masked images are resized to 512 x 512 matrix. Some sample masked X-ray images are shown in Figure 3.

The OC-SVM is quite similar to the standard SVM except that it aims to maximize the margin between positive and negative samples for better approximation of the class boundary. The OC-SVM has been widely used in identifying anomalies [36-38]. It uses the raw images or image features to learn and create a tight envelop around normal image data. When a new image is presented to the OC-SVM, it is accepted or rejected according to its resemblance to the training samples. The OC-SVM is useful in the situations where there are unbalanced classes in training data. For example, 99% of labelled images are from a single category and 1% of the labelled images are from other classes. In this case, the task is to identify the 1% outlier images that are not from the single category.

To train the OC-SVM, we used 502 normal X-ray images provided by Coal Services Health. The test image dataset is composed of normal and pneumoconiosis X-ray images from Wesley Medical Imaging, ILO, and NIOSH, including 15 pneumoconiosis and 26 normal X-ray images.

The kernel of the OC-SVM used is the non-linear radial basis function (RBF). There is an important adjustable parameter, 'nu', which has significant impact on final accuracy of classification. We have tested a range of values of this parameter, including 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, and 0.5. The test results show that the best sensitivity, specificity and accuracy are achieved when nu = 0.01 and nu = 0.5, respectively.

Table 1 Sensitivity, specificity and accuracy obtained using OC-SVM with the masked raw X-ray images as input

nu	Sensitivity	Specificity	Accuracy
0.01	0.4	**0.9231**	**0.7317**
0.5	**0.7333**	0.4615	0.5610

4.1.4 Experiments Using Isolation Forest with Raw Chest X-ray Images as Input

In these experiments, we used an Isolation Forest classifier [39] to identify the pneumoconiosis images. The Isolation Forest is derived from Random Forest (RF) [40, 41], and is adapted to isolate outliers in a dataset. The Random Forest is a broadly used learning method for classification, regression and other tasks by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes for classification problems or mean prediction for regression problems of the individual trees. The training and testing datasets used in the experiments are the same as that used in the experiments described in Section 4.1.3.

"Contamination Ratio" is an important parameter in Isolation Forest. This parameter has significant impact on the performance of the Isolation Forest. It is used to adjust the learned classification curve that encloses all normal data points more tightly or more loosely. The contamination parameter was chosen from a list of values [0.1, 0.2, 0.3, 0.4, 0.5], and our experiments show that the best accuracy, specificity, and sensitivity are achieved when the contamination = 0.1 and 0.5, respectively. The sensitivity, specificity and accuracy are listed in the following table.

Table 2 Sensitivity, specificity and accuracy obtained using Isolation Forest with raw chest X-rays as input

Contamination	Sensitivity	Specificity	Accuracy
0.1	0.3333	**0.8846**	**0.6829**
0.5	**0.9333**	0.2308	0.4878

4.1.5 Experiments using the Hybrid Model of Autoencoder and A Feed Forward Neural Network Classifier

The hybrid model of Autoencoder and the feed forward neural network combines the ability of deep neural networks to extract rich representation of image data with the one-class objective of creating a tight boundary around normal image data [42]. With the hybrid model, data representation in the hidden layer is customized for anomaly detection. This is different from other one-class classification approaches which employ a hybrid model of learning deep features using an Autoencoder and then feeding the features into a separate model for anomaly detection like One-Class SVM (OC-SVM).

Deep learning-based framework has been used as a standard feature extraction technique in image processing. In this study, we used an Autoencoder as our network architecture. The intuition behind this is that when the Autoencoder is trained to reconstruct the normal images only, then the anomalies cannot be fully recovered. Thus, the extracted features will exhibit discriminative characteristics, and will help improve the accuracy of the one-class classifier.

The Autoencoder contains three densely connected layers: input, hidden and output layers. The input layer takes 512 x 512 masked X-ray images as inputs, the size of the output layer is the same as that of the input layer. The output layer is used to reconstruct the input via the mean squared error loss. The size of the hidden layer is an adjustable parameter, and we have tested the following numbers, 16, 32, 64, 128, 256, aiming to get the best performance of classification. The Autoencoder is trained to obtain the representative features of the input images, then the

encoder layers of this pre-trained Autoencoder is copied and fed as input to the feed-forward neural network with one hidden layer. The weights of the encoder network are frozen while training the feed-forward neural network.

In addition, to better customize both the Autoencoder for the feature extraction and the feed forward neural network classifier to our dataset, we constructed an end-to-end architecture. Specifically, we added another two densely connected layers (one with 128 neurons and the other with one neuron) as the One-Class Neural Network (NN) classifier.

The training and testing datasets are the same as those used in the previous experiments. We adjusted two important parameters to tune our OC-NN model. One is the size of hidden layer in the Autoencoder part, and the other is the parameter 'nu', which has a similar function as the 'nu' in OC-SVM. Our experiments show that the best performance was achieved with the size of hidden layer 128, nu = 0.01, 0.1 and 0.5, as shown in the following table.

Table 3 Sensitivity, specificity and accuracy obtained from the hybrid model of Autoencoder and the feed forward Neural Network

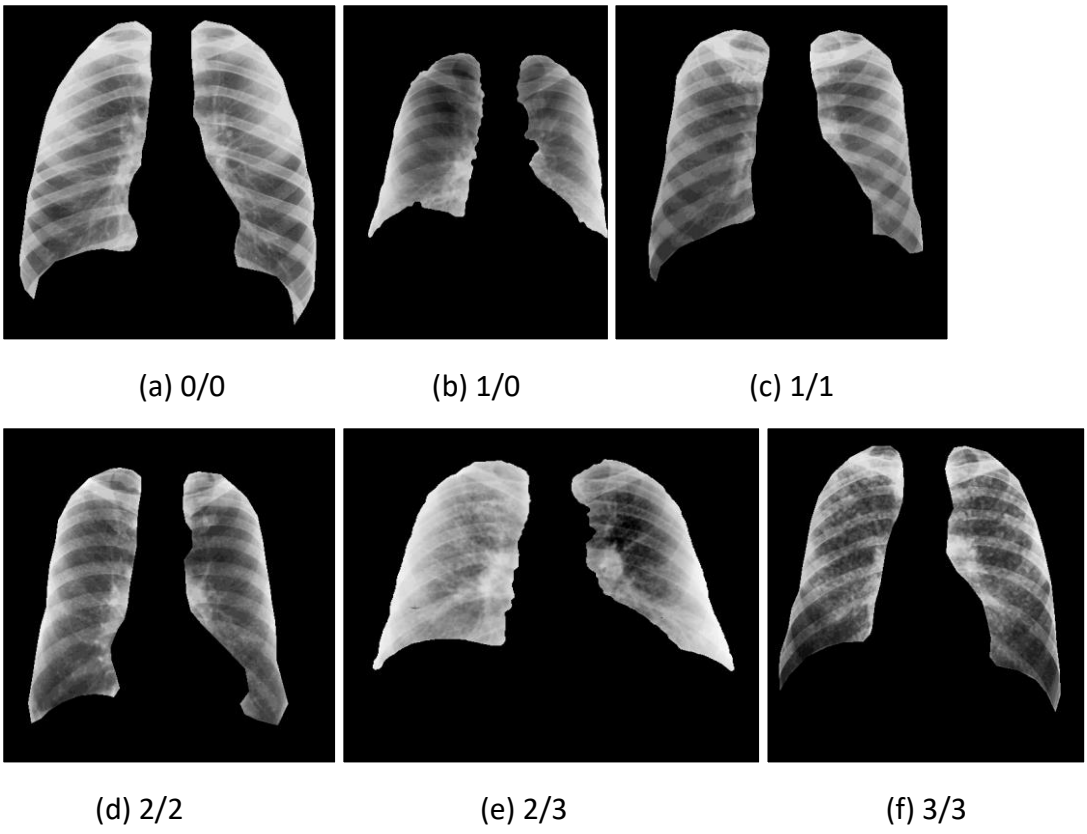
nu	Sensitivity	Specificity	Accuracy
0.01	0.0667	**0.8846**	0.5854
0.1	0.3333	0.7692	**0.6098**
0.5	**0.6**	0.1923	0.3415

4.2 Two-Class Classification Using Classical Machine Learning

We have also explored various two class classification methods to evaluate their performance. The two class classification methods are developed to distinguish between just two classes of objects. In this study, we employ the two class classification to identify normal or pneumoconiosis X-rays. Similar to the one-class classification experiments, we investigated several two class classification methods and their hybrid models.

4.2.1 KAZE and MLP Based Automated Detection of Pneumoconiosis

When the image dataset for training a machine learning model is not very large, classical machine learning tools are normally used. In this study, we have used the ILO Standard Radiographs, the X-ray images downloaded in B Reader Syllabus and the image dataset provided by Wesley Medical Imaging. The total number of images is 147, including 80 normal X-ray images and 67 pneumoconiosis images ranging from ILO 0/1 to 3/3. These images are split into two groups, 70% is used for training and 30% for testing. The lung fields of these images are segmented, and the lung field images are used for the training and testing. The following figure shows some examples of the masked lung field images used in the study.



(a) 0/0 (b) 1/0 (c) 1/1

(d) 2/2 (e) 2/3 (f) 3/3

Figure 6 Sample lung field images of various ILO categories used in the study

We used KAZE algorithm to extract local features from the images [43-44]. KAZE is a Japanese word meaning “wind”, it is defined as the flow of air on a large scale ruled by nonlinear processes. The algorithm is a novel multiscale 2D feature detection and description method in nonlinear scale spaces by means of nonlinear diffusion filtering. The evaluation reported in the paper [43] shows the KAZE outperforms the previous state-of-the-art methods in feature detection and description. Because the number of descriptors for different images varies, we cannot simply use the local features extracted as the input of a neural network. Instead, we turn the descriptors into a single histogram of visual words using the Bag of Words strategy. The histogram is then used as the input to our neural network.

The neural network we used is Multi-Layer Perceptron (MLP). Unlike many other models in Machine Learning that are constructed and trained at once. The MLP can be trained more than once, that is, the weights of the MLP can be adjusted based on the new training data when they become available. This is typically useful when we collect more images to improve the performance of the neural network.

From the 147 X-ray images, we randomly selected 56 normal X-ray images and 46 positive X-ray images for training the MLP and used 24 normal and 21 positive X-ray images which are not used in the training to test the MLP. The reading of the training images, KAZE feature extraction, creation of network, and training process took about 207.6 minutes, about 2.02 minutes per image. During the testing, to quickly calculate the histogram of visual words for each test image, a FLANN (Fast Library for Approximate Nearest Neighbours) model [45] was trained and employed. The reading of the test images and automated detection of pneumoconiosis took about 81.5

minutes, about 1.8 minutes per image. The confusion matrix for the automated detection of pneumoconiosis is shown below:

Table 4 Confusion matrix of the automated pneumoconiosis detection using the MLP

	Normal	**Positive**
Normal	18	6
Positive	**7**	**14**

The confusion matrix shows that the accuracy for the automated pneumoconiosis detection is 71.11%. With more images, this accuracy can be further improved.

4.2.2 Experiments Using SVM with Raw Chest X-Ray Images as Input

In these experiments, we used raw X-ray images as inputs of SVM for training and testing. We have tried different values of the parameter 'C' to observe the SVM classification results. The parameter 'C' is a regularization parameter that controls the trade-off of optimisation between the achieving of a low training error and a low testing error that is the ability to generalize the SVM classifier to unseen data. For large values of 'C', the optimization will choose a smaller-margin hyperplane if that hyperplane can get all the training images classified correctly. Conversely, a very small value of 'C' will lead to a larger-margin separating hyperplane, even if that hyperplane misclassifies more images.

In the experiments, we use the following setup:

- Training image dataset: 112 images including 56 normal and 56 pneumoconiosis X-ray images from Wesley Medical Imaging, ILO, NIOSH and Coal Services Health.
- Testing image dataset: 41 images including 26 normal and 15 pneumoconiosis X-ray images from Wesley Medical Imaging, ILO, NIOSH and Coal Services Health.
- We have included ILO 0/1 images in pneumoconiosis category for training and testing assuming they have some signs of pneumoconiosis that we did not want to miss.

The best performance of SVM is recorded when C = 1, and the experimental results are shown in the table below:

Table 5 Sensitivity, specificity and accuracy obtained using SVM

Sensitivity	Specificity	Accuracy
0.3333	**0.8846**	0.6829

4.2.3 Experiments Using the Hybrid Model of Autoencoder and SVM

With these experiments, we first trained the Autoencoder for feature extraction. The training dataset includes 502 normal X-ray images provided by Coal Services Health.

To train the two-class classifier SVM, as it requires a balanced positive-negative dataset, we carried out experiments with the same experimental setup in the previous section. Different values for the parameter 'C" have been evaluated, the best accuracy, sensitivity and specificity are observed when the hidden layer size is set to 128, and C = 1, 0.1 and 2, respectively. The experimental results are illustrated in the following table.

Table 6 Sensitivity, Specificity and accuracy obtained from the hybrid model of Autoencoder and SVM

C	Sensitivity	Specificity	Accuracy
1	0.7333	0.7308	**0.7317**
0.1	**0.9333**	0.3077	0.5366
2	0.5333	**0.7692**	0.6829

4.2.4 Automated Detection of Pneumoconiosis with Custom Image Features

We have quantified the profusion of opacities by, firstly, enhancing disk-like structures in a chest X-ray and, secondly, computing statistical features of the distribution of detected "disks" in each of the lung zones. We employed not one, but three disk-enhancing image filters, at different spatial resolution scales, in order to capture opacities of different sizes. Our implementation of the disk-enhancement filter is based on the published work in [46]. We restricted the search of disk-like structures to the lung fields only using pre-processed images as depicted in Figure 3 and Figure 6. Some examples of images where only circular structures are enhanced, and other structures are suppressed are given in Figures 7(a) and 7(c). One can visually appreciate differing amount of circular structures in a normal radiograph (Figure 7(a)) and a radiograph that has a 3/3 pneumoconiosis category.

Following this, we computed statistics of detected disks for each lung zone. The image with enhanced disk structures was binarized (see Figures 7(b) and 7(d)), and the four values that summarized the distribution of the disks in the lung zone were then computed, namely,

- *DotNo* – the number of disks in a zone
- *DotMean* - the average pixel intensity of disks in a zone
- *DotArea*- the average area of a disk, and
- *DotDensity*- the density of disks in the zone

The profusion of opacities in the *worst affected* zone is most important as it is the ILO grade of this zone which determines the ILO grade designated to the radiograph. We have chosen the worst affected zone in the lung based on the value of *DotDensity* feature as it describes the proportion of a lung zone that is covered with disk-like structures. The worst affected zone was selected at each spatial resolution scale. The resulting feature vector describing the profusion of small opacities in the radiograph consisted of 12 features: four statistical features for the worst affected zone selected at each spatial resolution scale.

As with KAZE features described in the previous section, with these custom features we used classical machine learning strategies to automatically detect normal chest x-rays and the ones with features of pneumoconiosis. We compared the performances of five different classifiers: Linear Support Vector Machine (SVM), Perceptron, K-Nearest Neighbour (KNN), Ridge Classifier and

Random Forest, using Leave One Out methodology: Each training set is created by taking all the samples except one, the test set being the sample left out. Thus, for n samples, we have n different training sets and n different test sets. This cross-validation procedure does not waste much data as only one sample is removed from the training set. The classification was very fast. It took between 0.3 sec and 2.2 sec for any of the five classifiers to perform Leave One Out on the 147 images. The performance of different classifier was measured in terms of accuracy, and the results are given in Table 7. For the best performing classifier, Ridge Classifier, we obtained sensitivity 63% and specificity 87% (which equals to false positive rate of 13%).

Table 7 Accuracy of different classifiers

Linear SVM	Perceptron	K-NN	Ridge Classifier	Random Forest
0.748	0.653	0.693	**0.769**	0.708

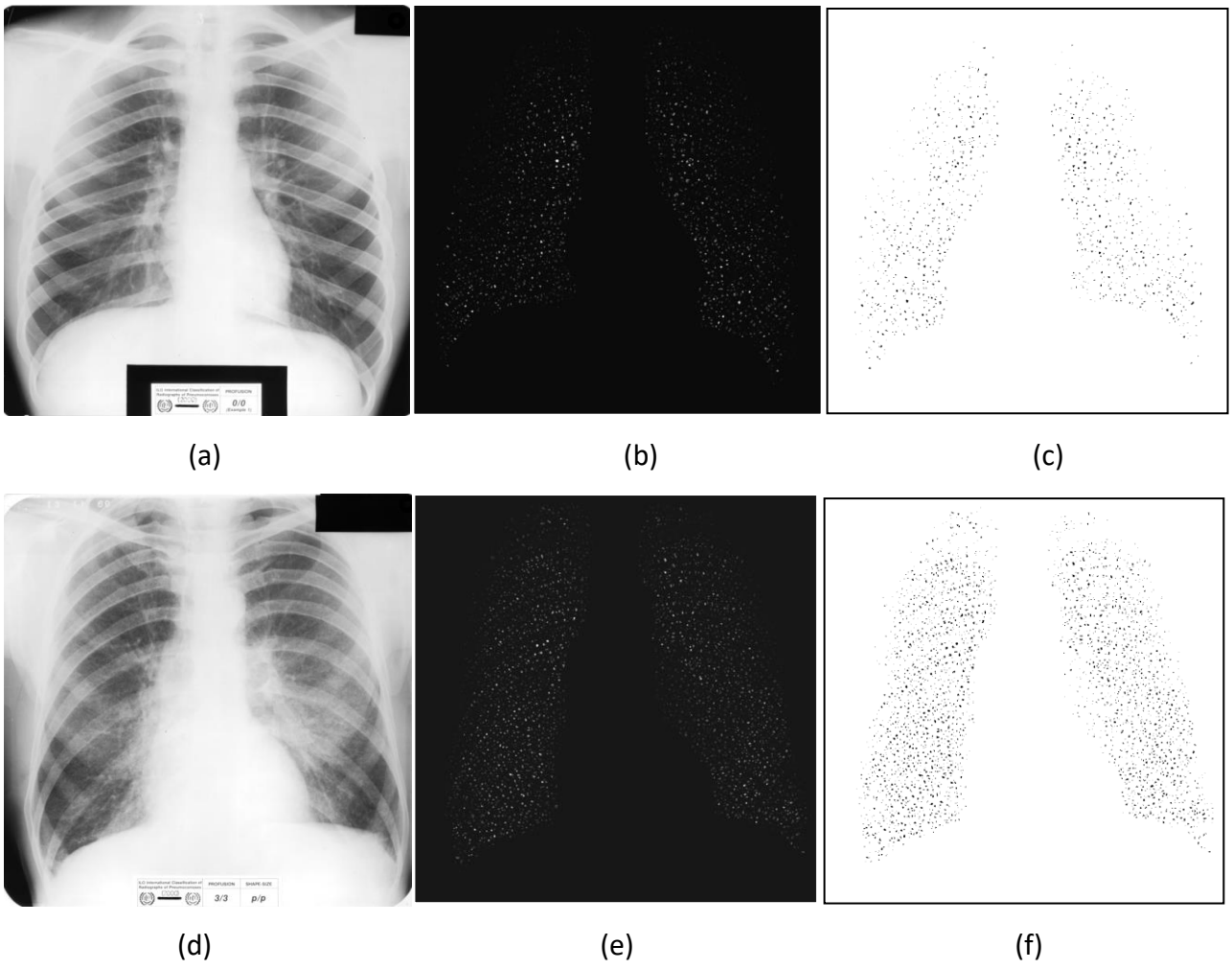


Figure 7 Examples of two radiographs filtered with a disk-enhancing filter, (a) normal radiograph, (d) radiographs with small opacities graded 3/3. In (b) and (e) disk-like structures are enhanced, and in (c) and (f) - separated from the background by thresholding the images.

We have demonstrated here that the custom image features devised specifically to quantitatively describe the radiographic features of pneumoconiosis are useful and show promising results in the automated classification. We have also shown that two linear classifier, Linear SVM and Ridge

Classifier, outperform other classifiers on this dataset. We believe that the overall classification performance, irrespective of a classifier or features used, will improve when more relevant data become available.

4.3 Classification of Profusion of Small Opacities

In this section we present a machine learning based method to classify pneumoconiosis into different categories with respect to the profusion of small opacities in the lungs (i.e. ILO grade). We used chest X-rays from previously described datasets. The results obtained with our method are presented in this section, however detailed discussions on the results as well as the limitations of our method are reported in Section 7.

4.3.1 Data Labelling

ILO Classification System [10] provides a set of Standard Radiographs that define four categories of the profusion of small opacities: 0, 1, 2, and 3. The ILO grade assigned to a radiograph is classified into one of 12 ordered subcategories: 0/-, 0/0, 0/1, 1/0, 1/1, 1/2, 2/1, 2/2, 2/3, 3/2, 3/3, 3/+. These subcategories are divisions of the continuum of increasing profusion of small opacities. The first number designates the ILO standard radiograph the patient radiograph most closely matches, and the second number reflects whether the patient radiograph could be considered to be between grades. Thus, the first number reflects the predominant profusion grade and the second number reflects whether the patient's radiographic profusion is slightly more or less than the standard radiograph of that ILO grade.

In our combined datasets, often only a few samples of some of the 12 categories are available, for example, there are only 4 images classified as 3/2 by experts. There is none classified as 0/- or 3/+. Our data are also very unbalanced – the number of normal images vastly exceeds the number of images in any other categories (Table 8). To make automated classification possible, we must aggregate available data into a smaller number of classes.

Table 8 Number of training images in each subcategory

Category	0/0	0/1	1/0	1/1	1/2	2/1	2/2	2/3	3/2	3/3
#Samples	82	3	8	14	4	4	12	8	4	8

One obvious option would be to have four classes - Normal (0), Class 1, Class 2, and Class 3 – based on the preferred category (the number on the left to the oblique stroke, for example, it is 3 in Subcategories 3/2, 3/3, and 3/+). In our case, such a division is still very unbalanced, where Normal class has 85 images while Class 3 only has 12 images. Therefore, we have decided to combine the subcategories into three broader classes that reflect the severity of profusion. To do that, we have considered the following three setups.

Table 9 Experimental setup 1

Class	0	1	2
Including subcategories	0/0	0/1, 1/0, 1/1, 1/2	2/1, 2/2, 2/3, 3/2, 3/3
Number of samples	82	29	36

Table 10 Experimental setup 2

Class	0	1	2
Including subcategories	0/0	0/1, 1/0, 1/1, 1/2, 2/1	2/2, 2/3, 3/2, 3/3
Number of samples	82	33	32

Table 11 Experimental setup 3

Class	0	1	2
Including subcategories	0/0, 0/1	1/0, 1/1, 1/2, 2/1	2/2, 2/3, 3/2, 3/3
Number of samples	85	30	32

In Experiments section we present results obtained with each experimental setup.

4.3.2 Methods

Previous work

In our previous work, we used custom image features, obtained per each lung zone and aggregated into a vector representing an entire lung, to classify an X-ray into normal or pneumoconiosis classes. The image features quantified the distribution of bright disk-like structures in a lung zone. The best classification performance achieved on the same dataset was 77% accuracy with 63% sensitivity and 87% specificity. A closer look at the image features revealed that a lot of normal disk-like structures inside the lung fields were enhanced even stronger than small opacities, for example, rib crossings and vessel cross-sections, therefore limiting the strength of such custom features to differentiate between normal structures and pneumoconiosis structures on an x-ray image, especially when image features are aggregated over large areas.

Proposed approach

In this report we propose to use local texture features extracted from small regions of interest (ROIs) placed inside the lungs, and a three-stage classification system, that is, firstly, classifies ROIs into one of three classes described in “Data labelling” section, and then derives a class for each lung zone from the results of ROIs’ classification. In the last stage, an image label is obtained by applying a fusion rule to the result of zone classification. We train separate classifiers for ROIs extracted from different lung zones.

Lung partitioning and ROI extraction

Lung fields were segmented from radiographs using multi-resolution pixel classification as described in Section 3. Each lung was automatically divided into three zones by dividing the vertical distance between the lung apices and the domes of the diaphragm into three equal parts and drawing a horizontal line at each division point. The algorithm assigns each zone a label, from 1 (Right Upper Zone, RUZ) to 6 (Left Lower Zone, LLZ), for an easy reference.

As shown in Figure 9, Regions of Interest were automatically fitted into the periphery of each zone. We opted to only cover the periphery of the lungs with ROIs because the part of the lungs closest to mediastinum, including the hilum, has a very complex textural appearance with multiple

overlapping structures, and might not be a very useful site for detection of small opacities, automatically or even manually.

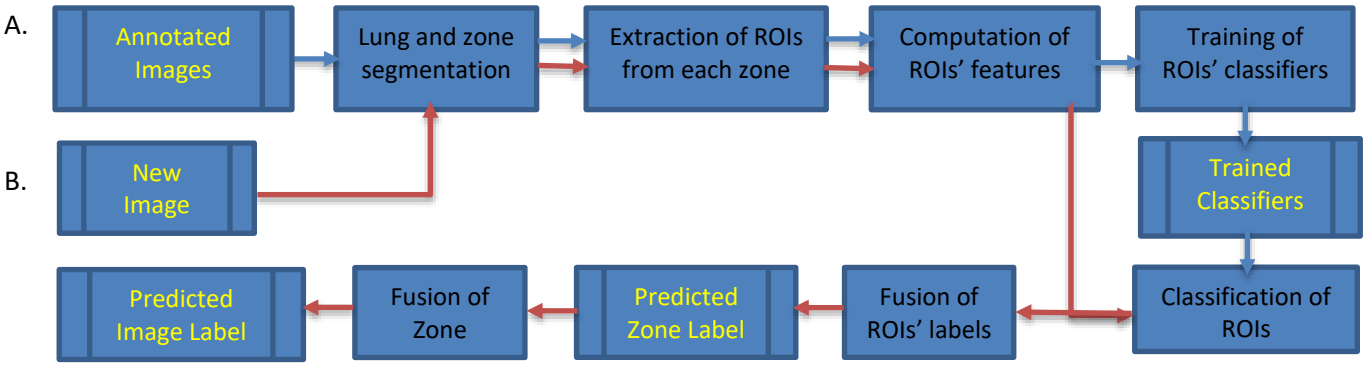


Figure 8 Flow chart of the proposed system. (A) Training phase (blue path), and (B) Testing phase (red path)

Local texture features

A powerful method for local texture analysis is filtering the image with a multiscale filter bank of Gaussian derivatives and calculating the moments of histograms from regions in the derived images. Using multiple scales allows us to characterize texture elements of different sizes, and analysis of local histogram considers the texture primitives regardless of their spatial distribution [47]. This general approach to texture characterization has been previously applied to detect interstitial abnormalities in chest radiographs and thoracic CT scans [48].

Following this approach in our project, radiographs were filtered with Gaussian derivatives of orders 0, 1, and 2 at five scales, s = 1, 2, 4, 8, 16. The four central moments - the mean, standard deviation, skewness and kurtosis – were computed for each ROI from the original and filtered images. Two position features were also added to the feature vector, namely, the x and y coordinates of the centre point of each ROI, computed relatively to the centre of mass of the lung containing it and scaled to the unit variance. In total, 126 features were extracted from each ROI.

ROIs' classification and label fusion

Six zone classifiers were trained with feature vectors extracted from ROIs. Each classifier only accepted features extracted from ROIs from a certain lung zone. Therefore, we built a RUZ classifier, an RMZ classifier etc. When a classifier is applied to a previously unseen test image, it assigns a class label to each ROI within its zone. To obtain a class label for the whole zone, the following fusion rule was applied: a zone is assigned to Class 0 (Normal), only if the number of ROIs classified as Class 0 is equal to or larger than the total number of ROIs classified as pneumoconiosis (Class 1 or Class 2). Otherwise, the zone is assigned to Class 1 if the number of ROIs classified as Class 1 is equal to or larger than the number of ROIs classified as Class 2. If the number of ROIs classified as Class 2 is larger, then the zone is assigned to Class 2. The left box in Figure 10 shows this rule in algorithmic notation.

To obtain a classification label for the whole image, the predicted zone labels were combined in the following way: if all six zones are classified as Class 0 (Normal), the image is also assigned to Class 0. Otherwise, the image is assigned to the same class as a prevalent zone label. See the right box in Figure 10 for the algorithmic notation of this rule.

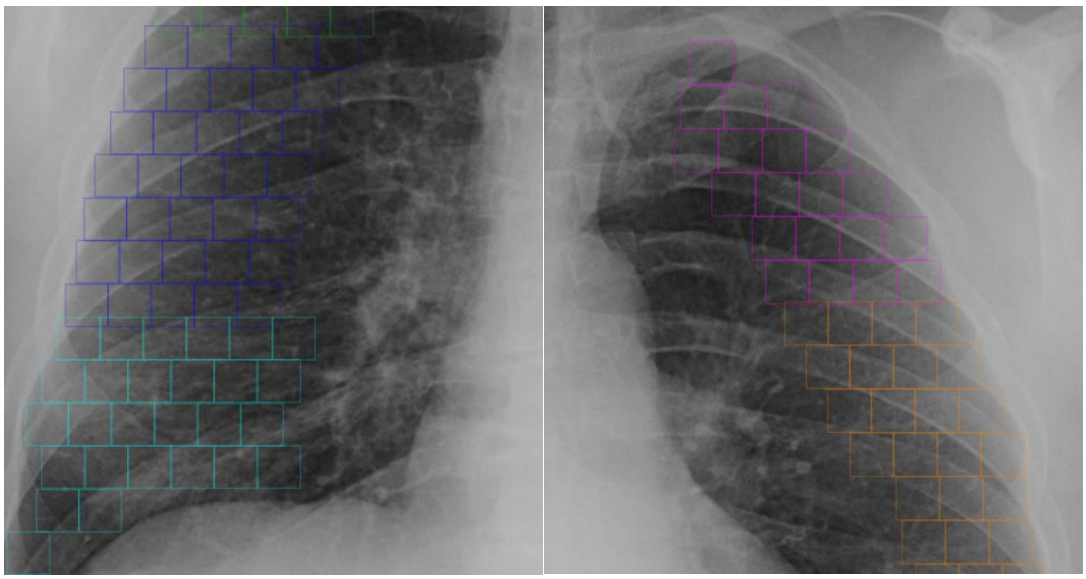


Figure 9 ROI coverage of the upper and middle zones of the left lung (A), and the middle and lower zones of the right lung (B).

```
R0 - ROIs classified as Class 0,
R1 - ROIs classified as Class 1,
R2 - ROIs classified as Class 2,
Lz - Assigned zone label

if R0 >= R1 + R2: Lz = 0,

else if R1 >= R2: Lz = 1,

     else: Lz = 2
```

```
Z0 - Zones assigned to Class 0,
Z1 - Zones assigned to Class 1,
Z2 - Zones assigned to Class 2,
L  - Assigned image label

if Z0 = 6: L = 0,

else if Z1 >= Z2: L = 1,

     else: L = 2
```

Figure 10 Label fusion algorithms: fusion of ROIs' label to obtain a zone label in the left box, and fusion of zones' labels to obtain an image label in the right box.

4.3.3 Experiments and Results

For practical considerations, images were downsized to the width 2048 pixels and an appropriate height to keep the same image aspect ratio. Pixel intensities in radiographs were normalized by histogram matching as described in Section 3.2. Radiographs from Wesley Medical Imaging and NIOSH B Reader Syllabus datasets were normalized, with radiographs from ILO Standard Radiographs serving as reference images.

We conducted image classification using three different data labelling setups for training, as described in "Data labelling" section. As in our previous work, Leave One Out methodology was used: each training set is created by taking all the images except one, the test set being the image left out. A variety of well-known classifiers were employed in the ROI classification stage, to find the most successful ones. In the tables below the results obtained with the best performing Multi-layer Perceptron (MLP) Classifier are presented for each data labelling setup.

Alongside the performance measures for three-class classification, such as a confusion matrix, precision, recall, and F-score for each class (Tables 12 - 14), we also computed performance measures for binary classification (Table 15). To convert three-class classification results into binary classification results, we used the following rule:

$TN = X_{00}$

$TP = X_{11} + X_{22} + X_{21} + X_{12}$

$FN = X_{10} + X_{20}$

$FP = X_{01} + X_{02}$

In this rule, X_{ij} is the number of images belonging to Class *i* (true class) and classified as Class *j* (predicted class).

Table 12 Three-class classification results for Setup 1

	Class 0	Class 1	Class 2
Predicted class 0	**67**	9	1
Predicted class 1	5	**11**	3
Predicted class 2	10	9	**32**
Precision	0.870	0.579	0.627
Recall	0.817	0.379	0.889
F-score	0.843	0.458	0.736

Table 13 Three-class classification results for Setup 2

	Class 0	Class 1	Class 2
Predicted class 0	**65**	12	0
Predicted class 1	9	**13**	4
Predicted class 2	8	8	**28**
Precision	0.844	0.5	0.636
Recall	0.793	0.394	0.875
F-score	0.818	0.441	0.737

Table 14 Three-class classification results for Setup 3

	Class 0	Class 1	Class 2
Predicted class 0	**73**	11	1
Predicted class 1	4	**11**	2
Predicted class 2	8	8	**29**
Precision	0.859	0.647	0.644
Recall	0.859	0.367	0.907
F-score	0.859	0.468	0.753

The F-score is the harmonic average of the precision and recall, and, when computed for the three different setups, it allows us to compare them among themselves. A F-score value ranges between 0 (worst) and 1 (best). We prefer F-score to accuracy, as a performance metric, because we are

dealing with unbalanced classes, and the overall accuracy (for example, accuracy = 0.769 for Setup 3) doesn't reflect how badly the classifier performs for Class 1. From the results presented in Tables 12-14, we see that F-score is the highest for Setup 3 for each of the three classes. We can also see that correctly classifying radiographs belonging to Class 1 is the hardest (F-score is 0.468 in Setup 3). We have also applied a χ^2 test to the confusion matrices in Tables 12 - 14 and found that the confusion matrices for Setup 2 and Setup 3 are significantly different ($p < 0.05$).

We converted the results above into binary classification results and computed the accuracy, sensitivity (recall), specificity, precision and F-score for each setup. Note that recall and sensitivity are the different names for the same measurement, true positive rate: *TP / (TP + FN)*. Specificity measures the true negative rate: *TN / (TN + FP)*, while precision measures positive predictive value: *TP / (TP + FP)*.

Table 15 Binary classification results

	Setup 1	Setup 2	Setup 3
Accuracy	0.830	0.801	0.836
Sensitivity (recall)	**0.846**	0.815	0.806
Specificity	0.817	0.793	0.859
Precision	0.786	0.757	0.806
F-score	**0.815**	0.785	0.806

The results in Table 15 allow us to compare the performance metrics for the three setups. Although a χ^2 test applied to binary confusion matrices did not show a statistically significant difference among them, the higher sensitivity (recall) and F-score for Setup 1 indicates that this is a preferred way of labelling data in the absence of a more representative dataset. Obviously, sensitivity is an important metric as a cost of misclassifying a positive (pneumoconiosis) image is higher than erroneously classifying a normal radiograph as one with abnormalities.

5 Deep Learning Based Automated Pneumoconiosis Detection

Deep learning has become very popular and has been used pragmatically in many industry domains. However, one common barrier for deep learning to solve real-world problems remains the amount of labelled training data. In practice, imbalanced datasets often come up with majority of training data from a single class and a limited number of training samples from other classes. This can lead to biased prediction in favour of the majority class.

In this section, we report our experimental results using various deep learning schemes for the detection of pneumoconiosis.

5.1 Automated Pneumoconiosis Detection on Chest X-Rays Using Cascade Learning with Real and Synthetic Radiographs

For pneumoconiosis detection, we had abundant training data for normal X-rays; however, the number of X-rays with features of pneumoconiosis was limited. To address this issue, we propose a cascade learning architecture for the automated pneumoconiosis detection. The following figure shows the proposed cascade learning architecture, which is further detailed in the following sections.

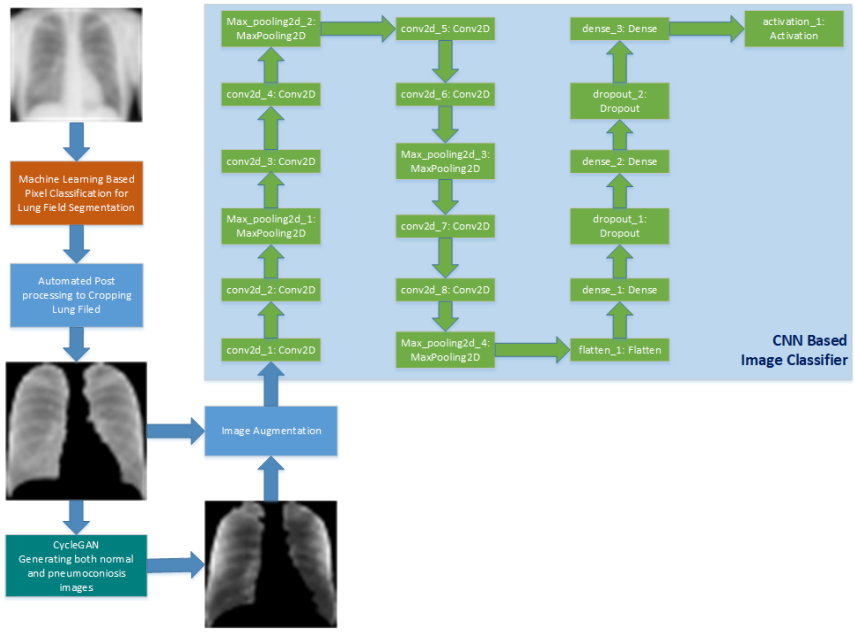


Figure 11 The overall architecture of the proposed cascade learning model

5.1.1 CycleGAN Image Generator

CycleGAN was proposed to capture special characteristics of one image collection and translate the characteristics into the other image collection [16]. It can be used to do image-to-image

translation and leverage the imbalanced training datasets. In this work, we train a CycleGAN using our 56 normal and 56 pneumoconiosis images to generate 1,000 normal and 1,000 pneumoconiosis images, respectively. Experiments show that overall good accuracy is achieved when using the synthetic images generated by CycleGAN trained for 30 epochs.

5.1.2 CNN Based Image Classifier

The input of our CNN based image classifier are images of 256 x 256 in dimension. The classifier is trained to classify an image into the category of either normal or pneumoconiosis.

The CNN model is composed of 15 layers as shown in Figure 11. It includes 8 convolutional layers to extract feature maps. We start with 32 filters to extract low-level features, and double the number of filters to 64, then 128 and 256 to detect high-level detailed features. The kernel size used for these filters is 3 x 3 and stride is 1 x 1. The activation function used is ReLU. Four pooling layers are employed to down sample the feature maps and provide spatial variance. There are also three dense layers with all input nodes of each dense layer connected with all nodes of its next layer. Dropout is used in the first two dense layers to prevent overfitting. The last layer of the classifier uses sigmoid activation function and outputs a probability score for each class – normal and pneumoconiosis.

For the classifier, its input is a chest X-ray image *X* and the output is a binary label $y \in \{0, 1\}$ representing the absence or presence of pneumoconiosis, respectively. During the training, we use binary cross-entropy as loss function, and RMSprop optimizer. We optimize the binary cross entropy loss:

$$L(\hat{y}, y) = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)] \quad (2)$$

Where $L(\hat{y}, y)$ is the binary cross loss, y_i is the true value (0 or 1 for binary classification) and $\widehat{y_i}$ is the predicted probability of the label y_i, and N is the number of training samples.

5.1.3 Image Augmentation

All images including training, validation and testing samples are normalized so that their pixel values are between 0 and 1. For the training images, their mean is set to 0 by subtracting the mean value of the training dataset from each training image. Each training image is also divided by the standard deviation of the training dataset. To increase the diversity of the training dataset, the training images are randomly zoomed with a range of 0.9 to 1.1, and flipped horizontally, and their pixel intensities are sheared with an angle of 0.01 degrees. Apart from scaling the intensities to the range of [0, 1], no other augmentation was done for the validation and testing images.

5.2 Automated Pneumoconiosis Detection on Chest X-Rays Using Transfer Learning with Local Texture Patches

We have also considered a different approach to address the limited number of pneumoconiosis X-rays. In this approach, instead of generating new artificial images we parsed existing images into small regions of interest thereby increasing the number of available samples. The flow chart in Figure 8, Section 4.3, illustrates this classification model. The differences from the method

presented in Section 4.3 are in labelling of local patches, feature extraction and image classifiers used, and how the fusion of classification results is performed to obtain an image label.

5.2.1 Data Labelling

Local patches, or regions of interest (ROIs) were extracted from the lung periphery in a similar fashion as displayed in Figure 9, however, this time we have also considered larger and overlapping ROIs. The detailed settings will be described in Section 5.3. Since we indent to perform a binary classification of radiographs, each ROI has been labelled as Normal or Pneumoconiosis, in accordance with the label of a zone a ROI has been extracted from. For 147 out of 153 radiographs in our dataset, zone labels were provided, and for the six remaining (abnormal) images we assumed that each zone was abnormal.

5.2.2 Classifying Local Patches with DenseNet Image Classifier

For each zone we have trained a separate DenseNet – a densely connected convolutional network [14], which is a popular deep learning architecture in computer vision and has been previously used with CheXNet [13] – a tool that successfully detects pneumonia in chest radiographs. For our purposes, we have slightly modified the 121-layer DenseNet architecture by replacing a multi-class prediction layer with a two-class prediction layer and a binary cross-entropy loss function (see Eq. (2)). The input of the classifier is a ROI – a small square image within a lung field. No custom features are extracted from it. The DenseNet classifier is initialized with pre-trained weights obtained with the popular ImageNet database [50], and further trained with ROIs extracted from our training radiographs. ROIs extracted from normal zones of radiographs with pneumoconiosis are not used. All ROIs are normalized so that their pixel values are set between 0 and 1, and, additionally, ROIs used for training, are flipped, to increase diversity in the dataset. The output of the classifier is a probability between 0 and 1 that a ROI has features of pneumoconiosis.

5.2.3 Classifying Images

It is expected that different ROIs from the same lung zone might receive a wide range of probabilities of having features of pneumoconiosis: firstly, there will be classification errors (especially considering we did not have true labels for each ROI to train the classifier), and, secondly, there are likely to be ROIs within a pneumoconiosis affected zone that do not have features of pneumoconiosis, and, vice versa, a normal zone could contains ROIs with features resembling pneumoconiosis. However, it is reasonable to assume that we see more ROIs with higher probabilities of pneumoconiosis in pneumoconiosis-affected zones, and, similarly, less ROIs with higher probabilities of pneumoconiosis in normal lung zones.

Therefore, we derive a zone's probability of having pneumoconiosis as a *p-th percentile* of its ROIs probabilities of having pneumoconiosis. Let's denote a zone's probability as P_z, where z is one of [1, 2, 3, 4, 5, 6]. To obtain an image probability of being normal or having pneumoconiosis, we need to further aggregate zonal scores into one result. To do so, we reason that an image has features of pneumoconiosis if *any* of the zones has features of pneumoconiosis. Thus, we multiply the probabilities that the zones are normal. If every zone has a high probability of being normal, the resulting probability of the image being normal is also high. If any zone has a high probability

of being abnormal (which is the same as a low probability of being normal), the resulting probability of the image being normal will diminish. Such an aggregator was suggested in [51] for the task of detecting chest radiographs with tuberculosis.

The weighted probability of an image having features of pneumoconiosis is given by the following equation:

$$P = 1 - \prod_{z=1}^{6}(1 - w_z P_z) \quad (3)$$

where *z* is a zone number, and w_z is a weight for each zone. We weigh each zone's probability based on a classification performance for this zone. As an indicator of the zone classification performance, the Area Under ROC curve (AUC) is used. If AUC is below some threshold T, this zone is not considered, and if AUC = 1, it is fully taken into account. Therefore, w_z is computed as follows:

$$w_z = \max\left(\frac{AUC_z - T}{1 - T}, 0\right) \quad (4)$$

5.3 Experiments

In this section, we report our experimental results using the proposed cascade learning model and compare the results from various popular machine learning models we have evaluated, including ROI-based transfer learning model described in Section 5.2.

5.3.1 Experimental Setup

Among the coal mine worker chest X-ray datasets we collected, there are an abundance of normal X-rays and only 71 pneumoconiosis images. We set aside 56 pneumoconiosis images (80%) for training and 15 images (20%) for testing as described in Section 4.2.2. To use the same number of images from different classes for training, we set aside 56 normal images for training and 26 for testing. The abundant X-ray images from NIH are used for ChexNet based transfer learning. More details can be found in the following sections.

5.3.2 Experiments using CheXNet Based Transfer Learning

CheXNet is developed by Stanford Machine Learning Group to detect pneumonia from chest X-rays [13]. It is a 121-layer dense convolutional neural network trained on ChestX-ray14 image database, containing over 100,000 X-ray images with 14 diseases. In the experiments, we used a CheXNet model pre-trained with the ChestX-ray14 database as a starting point, and retrained it using 1,056 normal and 1,056 pneumoconiosis images, respectively. The 1,056 training images for each class include 1,000 synthetic images generated by CycleGAN and 56 real X-ray images. The CycleGAN was trained on the 56 normal and 56 pneumoconiosis images. The classification results for the testing dataset are demonstrated in Table 17.

5.3.3 Experiments with the Proposed Cascade Learning Model

Experiment 1 – Using Both Lung Fields in a Single X-Ray Image

To evaluate our proposed model, we used the same training and testing datasets as used for retraining the pre-trained CheXNet above. The testing dataset includes 41 images (26 normal and 15 pneumoconiosis images), and the training dataset has 1,056 normal and 1,056 pneumoconiosis images. The 1,056 training images for each class include 1,000 images generated by CycleGAN and 56 real X-rays. For each class, we split the 1,056 images into two datasets with 792 for training (75%), and 264 for validation (25%). The following figure shows the original lung mask (left) and CycleGAN generated image (right) for training.

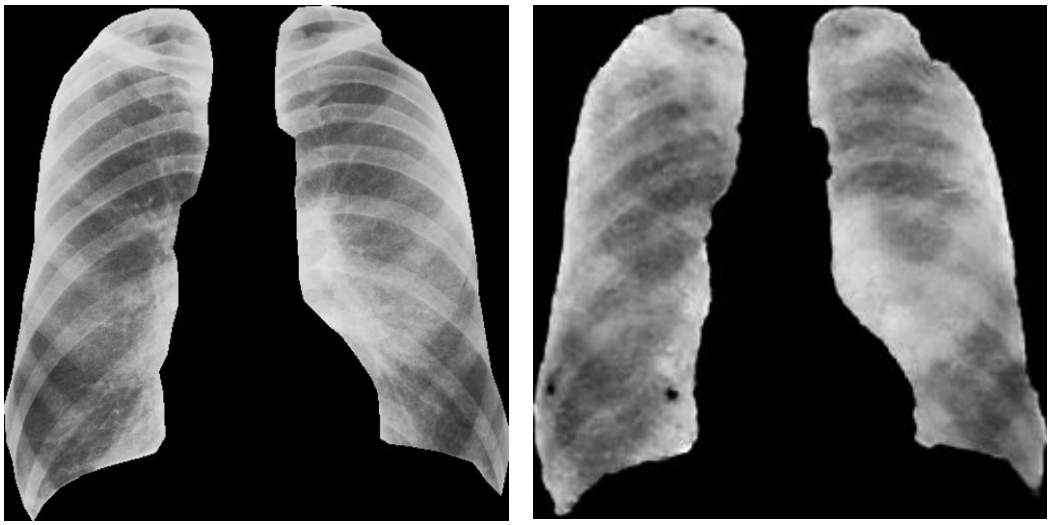


Figure 12 Training images - an original lung field X-ray image (left), and a CycleGAN generated X-ray image (right)

Training of the Image Classifier

For the training, we used the following hyper parameters: Learning Rate = 0.0001, Epochs = 20, Batch Size = 32. The dimension of the training images is 256 x 256. The training was conducted on a GPU workstation with an Intel 18-Core i9 2.6 GHz CPU, 128GB RAM, and 4 Titan Xp GPUs. The training for 20 epochs took only 6 minutes 52 seconds. During the training, the log-loss for the training images was between 0.058 and 0.691, and 0.058 at the end of the training; for the validation images it was between 0.029 and 0.691, and 0.052 at the end of the training. The classification accuracy for the training data was between 53.5% and 98.3%, and 98.3% at the end of the training; for the validation data it was between 50.2% and 99.6%, and 99.4% at the end of the training. The following figure shows the log-loss and accuracy during the training.

Testing of the Image Classifier

After the training, we tested our model with the test dataset. Only one pneumoconiosis X-ray image and 3 normal X-rays were misclassified. The overall classification accuracy is 90.24%, the sensitivity is 93.33% and the specificity is 88.46%.

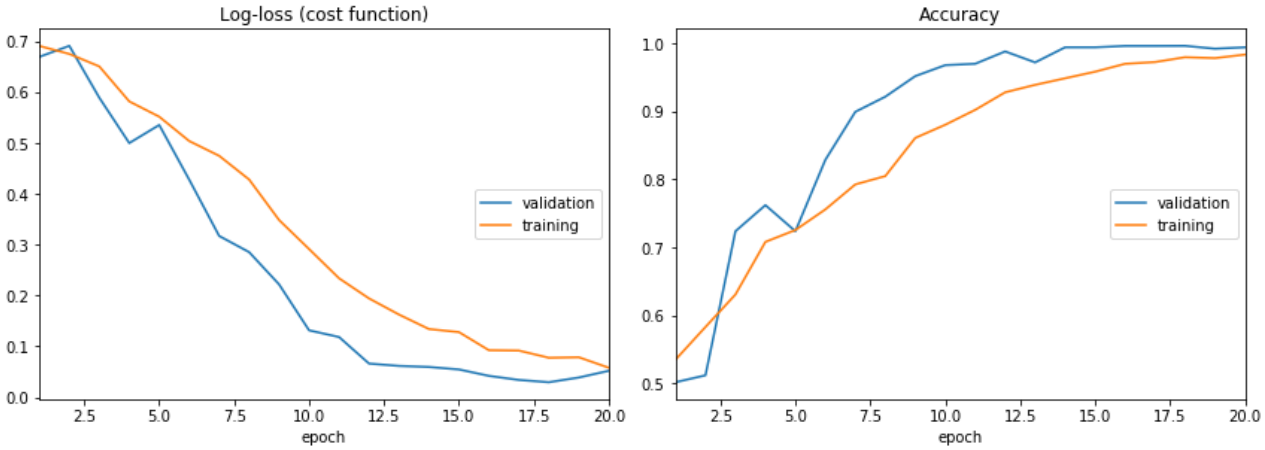
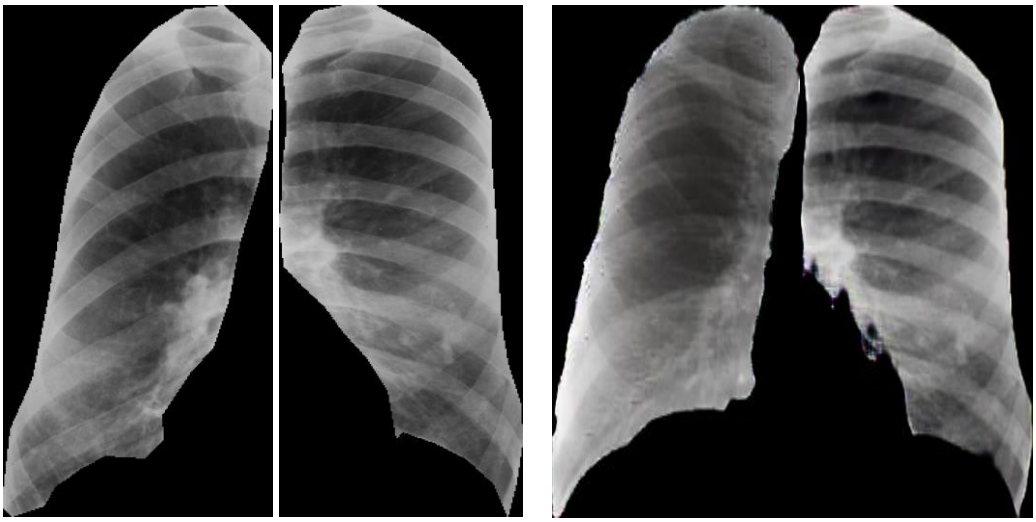


Figure 13 Loss and accuracy during training and validation

Experiment 2 – Using Left and Right Lung Fields Separately as Two Images

We have also conducted experiments by splitting left and right lung fields of each X-ray image into two images as shown in the figure below: (a) is the original right lung field image, (b) is the original left lung field image, (c) is CycleGAN generated right lung field image, and (d) is CycleGAN generated left lung field image.



(a) (b) (c) (d)

Figure 14 The original and CycleGAN generated images: (a) the original right lung filed image; (b) the original left lung filed image; (c) CycleGAN generated right lung filed image; and (d) CycleGAN generated left lung filed image

The testing dataset includes 82 images (54 normal and 28 pneumoconiosis images), and the training dataset has 2,112 normal and 2,112 pneumoconiosis images, including 1,000 left and 1,000 right synthetic lung images generated by CycleGAN and 112 real X-rays for normal and pneumoconiosis classes, respectively. For each class, we split the 2,112 images into two datasets with 1,584 for training (75%), and 528 for validation (25%).

The same data augmentation methods are applied, and the same training parameters are used as in Experiment 1. Figure 15 and Figure 16 demonstrates the classification accuracies and losses for

training and validation datasets during the training for Experiment 2. The final experimental results are shown in Table 17 for comparison.

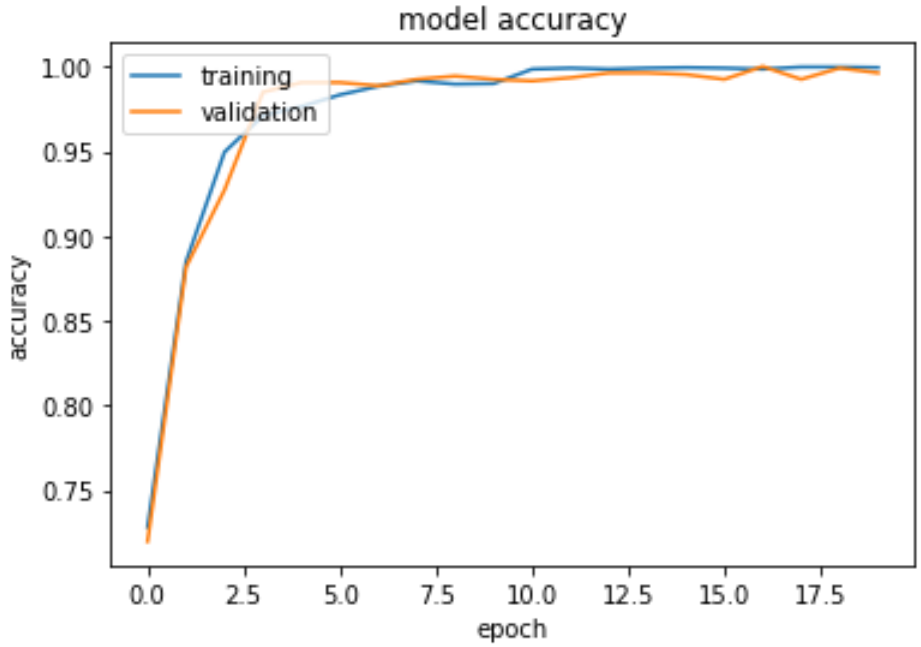


Figure 15 Classification accuracies for training and validation datasets during the training for Experiment 2

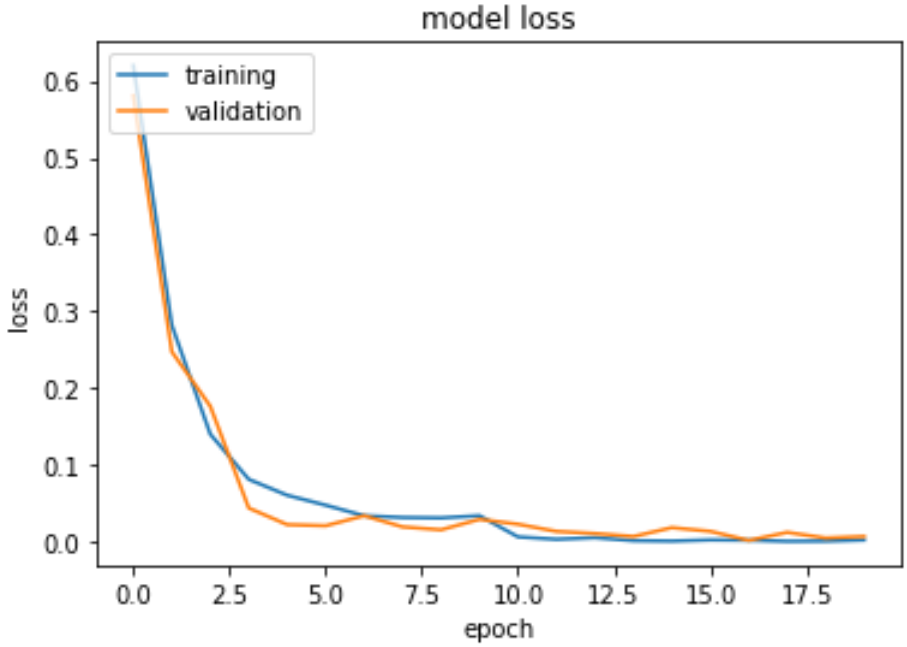


Figure 16 Losses for training and validation datasets during the training for Experiment 2

5.3.4 Experiments using ROI-based Transfer Learning Model

For these experiments, we extracted three types of ROIs from the lung periphery: 64 x 64 non-overlapping ROIs, 64 x 64 overlapping ROIs and 96x96 overlapping ROIs. The overlapping ROIs had

50% of overlap in horizontal and vertical directions. For training, we used the following parameters: the batch size of 32, the number of epochs of 50, and the initial learning rate of 0.001. The learning rate was reduced by 10% every time when the loss hit a plateau. Our best results were obtained with 64 x 64 overlapping ROIs. We selected the 75th percentile on ROIs' probabilities to characterize a zone's probability of being abnormal and computed the AUC for each lung zone using the same 41 test images as described in Section 5.3.1. The classification results per zone are presented in the following Table:

Table 16 AUC values for each lung zone. The closer AUC is to 1, the better a classifier distinguishes between the two classes of data

	RUZ	RMZ	RLZ	LUZ	LMZ	LLZ
Area Under Curve	0.94	0.90	0.85	0.86	0.93	0.91

Applying Eq. 3 and 4, with threshold T = 0.8 given that all the zones had AUC values higher than 0.8, the AUC = 0.92 is obtained for the test dataset. Figure 17 shows the corresponding ROC curve. By selecting an appropriate point on the curve that allows for the best sensitivity and specificity, we obtain 93.33% sensitivity, 80.77% specificity, and 85.37% overall accuracy.

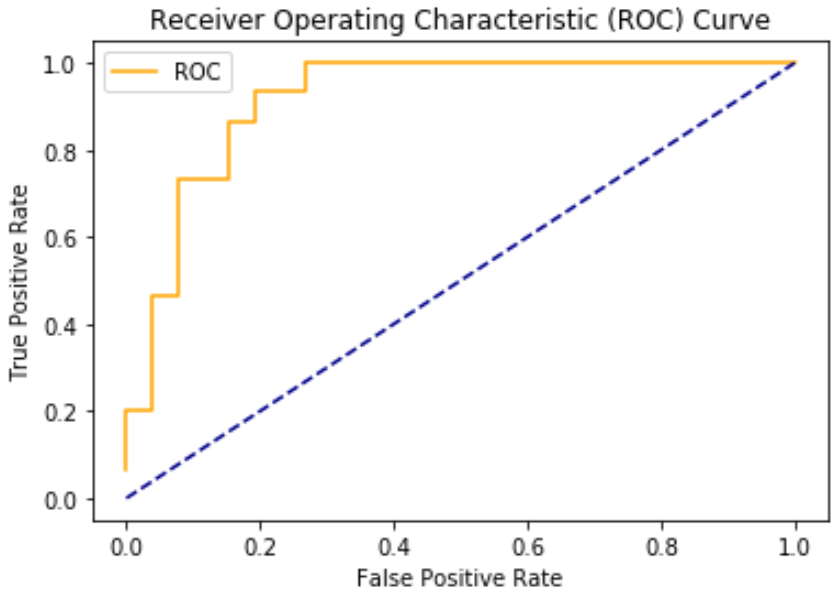


Figure 17 A ROC curve for the test dataset

Comparison of the Classification Results from Different Machine Learning Model

The table below compares the results from our models and other machine learning algorithms we evaluated. It clearly shows the proposed cascade learning model outperforms the others.

Table 17 Comparison of pneumoconiosis detection results obtained from different machine learning models

Method	Sensitivity	Specificity	Overall Accuracy
CheXNet Based Transfer Learning	73.33%	80.77%	78.05%

Proposed Cascade Learning (Experiment 1)	**93.33%**	**88.46%**	**90.24%**
Proposed Cascade Learning (Experiment 2)	90.74%	89.29%	90.24%
ROI-based Learning	93.33%	80.77%	85.37%

6 Black Lung Prediction Demo

In this section a web application is described that demonstrates our best performing automated pneumoconiosis prediction tool. This tool is utilizing the cascade learning approach described in Section 5 to predict whether a chest radiograph is normal or has features of pneumoconiosis. A user only needs a web browser to access this tool. We describe the implementation and web interface in the following subsections.

6.1 Web Interface

Black lung prediction demo can be found at http://confederate.csiro.au/. A page layout is straightforward as shown in Figure 18. Upon opening this web page, a user is presented with thumbnail images of 12 normal chest radiographs and 12 radiographs with pneumoconiosis. More images are revealed by clicking “Show more images” button. In total, 26 normal radiographs and 15 radiographs with pneumoconiosis are available on the web page. These are the same test images as in Section 5.3.1, providing the user with an opportunity to validate our classification algorithm.

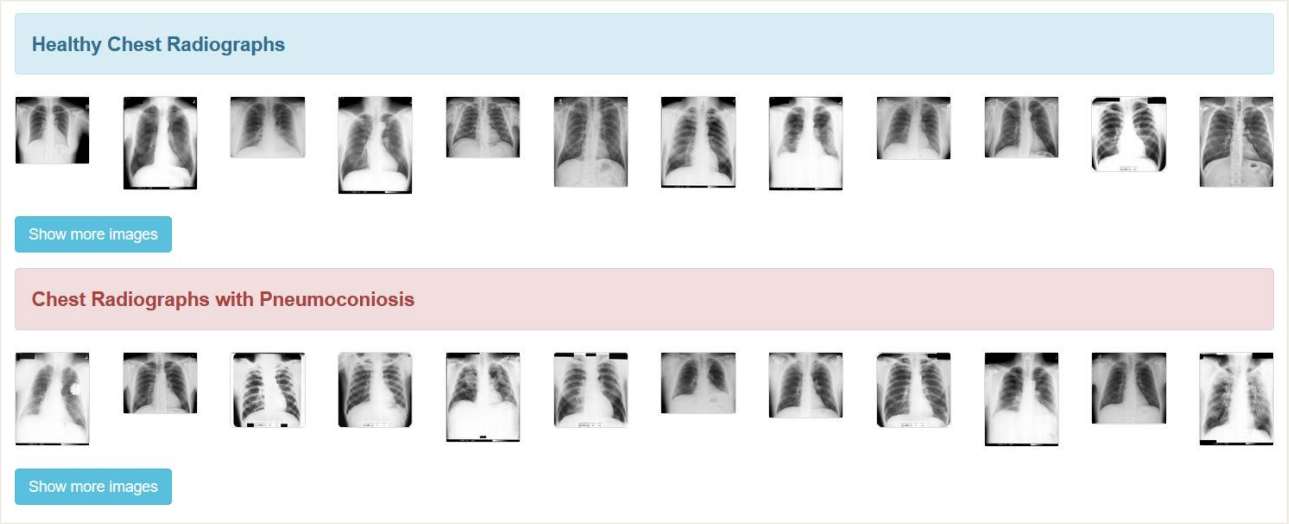


Figure 18 A user is presented with a choice of test radiographs.

Each small image is clickable: a larger view of a selected radiograph appears at the bottom of the page, where its classification process can be launched with “Predict” button as shown in Figure 19. After a few seconds, a prediction for the radiograph appears next to “Results:” label, together with the information whether it was a correct or incorrect prediction. Figure 20 and**Error! Reference source not found.** Figure 21 demonstrate examples of correct and incorrect classification results.

It is also possible to view a radiograph in its original resolution by clicking on the image. A new tab with the full-sized radiograph opens in the browser. Depending on a user’s screen size, they might need to manipulate a browser’s zoom function to enlarge the radiograph.

Now, the web demo only operates on 41 pre-selected images, but we might implement an upload function in the future.

Figure 19 A larger chest X-ray. By clicking "Predict" button a user starts an image classification algorithm for the displayed image

Figure 20 An example of correct prediction

Figure 21 An example of incorrect prediction

6.2 Implementation

In the back end we used a pre-trained deep learning model previously described in Section 5.1. The model and its best training weights were converted to a tensorflow.js model that we loaded into tensorflow.js - a JavaScript library for training and deployment of machine learning models in the browser. In this demo, we only used the algorithm to obtained predictions for test images, as the model had been already trained offline (Section 5.3.3). It only takes a few seconds to compute a prediction for the first image, and around one second per image for subsequent images, on an average PC. The demo runs in popular browsers such as Firefox, Chrome and Microsoft Edge (in Windows 10).

7 Summary and Discussion

Pneumoconiosis is incurable, prevention is the key to management. Early detection of pneumoconiosis through routine health screening is critical to preventing progression of disease, and complications including chronic disablement and death. Until the present day, there has been a lack of systematic, automated, and objective systems for detecting the presence of pneumoconiosis and assessing its progression in individual coal miners other than by expert radiologists. The insensitivity of chest radiographs for the detection of early pneumoconiosis, the inter- and intra-reader variability in interpreting a chest X-ray, and the shortage of B-readers each contribute to difficulties in identifying these occupational diseases.

We have developed a cascade machine learning algorithm which automatically detects pneumoconiosis from chest X-rays. This method employs a convolution neural network for image classification and utilizes a generative adversarial network to generate synthetic chest X-rays to train the image classifier. The proposed method outperforms others and achieves a sensitivity of 93.33%, a specificity of 88.46% and an overall accuracy of 90.24%. We hope this technology can be potentially used for the pre-screening of occupational lung diseases, and to address the issues of variability in identifying pneumoconiosis, and the shortage of B-readers. The cascade learning model can be potentially used in other medical imaging applications when the training dataset is imbalanced or lacks diversity.

To prepare the image data for training the machine learning models used in this study, we have developed and implemented algorithms for automated lung field and zone segmentation for both digitised analogue X-ray images and digital radiographs. The algorithms have been applied to the lung segmentation in the study.

With the methods based on statistical image analysis, we focused our effort on employing local texture features, i.e. texture features extracted from relatively small local patches, to achieve global classification results. Such features are potentially very informative in characterising ill-defined diffuse abnormal changes in a local textural appearance of the lungs and have been employed previously in published works, such as [47, 48, 49].

The datasets that were available to us contained so called weakly labelled data, meaning that the exact locations of abnormalities in training data were unknown. This is quite common since obtaining manual delineations of diffuse opacities is laborious and likely to produce unreliable ground truth. However, for most of our datasets zone-based ground truth was available, which allowed us to extend the true labels of zones to patches extracted from a corresponding zone. We limited the extraction of patches to a periphery of each zone in order to reduce the amount of overlap of different structures projected onto a 2D image of the chest.

We used the features extracted from local patches and the true labels of these patches to train six classifiers. Each of these six classifiers is for the patches extracted from one of the six lung zones. Once the classifiers were trained, they could be applied to patches from a new (previously unseen) image to assign each patch a predicted class label. Next, we combined patch labels to obtain a zone label. Unless all zones were labelled as normal, the image was assigned to a class that had

most zone labels. We evaluated our results using a three-class data labelling setup, as well as within a binary classification setup, where we combined all abnormal classes into one. The latter allowed us to compare this local classification approach to the other two-class classification methods described in this report.

With the classical machine learning based methods, we have investigated several approaches for the identification of pneumoconiosis on chest X-rays. The one-class classification is designed for imbalanced data in which one of the classes significantly outnumbers other classes. The real-world datasets are often predominately composed of normal examples with only a small portion of abnormal cases, like the normal X-rays vs pneumoconiosis X-rays. We have explored Autoencoder, SVM, Isolation Forest, Feed Forward Neural Networks and their hybrid models, and our experimental results show that the best sensitivity (93.33%) was observed when using Isolation Forest with raw chest X-ray images as input. Our experiments have also demonstrated that the best specificity (92.31%) and accuracy (73.17%) were obtained when using One-Class Support Vector Machines (OC-SVM) with raw chest X-ray images as input.

We have also investigated classical machine learning based two-class classification to identify normal and pneumoconiosis X-rays. By mixing the X-ray image data acquired from multiple sources, we managed to get an equal number of training images from each class. Several machine learning models have been trained and compared, including SVM, Autoencoder, MLP, Perception, K-NN, Ridge Classifier, Random Forest and some of their hybrid models. To mitigate the shortage of pneumoconiosis X-ray images, transfer learning has been employed by using pre-trained machine learning models for deep feature extraction and using the deep features to train classifiers to detect the pneumoconiosis X-rays. Our experimental results show that the best sensitivity (93.33%) was observed when using the hybrid model of Autoencoder and SVM, however the specificity for this method was low. Our experiments also show that most cases of misclassification are between X-rays with ILO grades of 0 and 1.

Both one-class and two-class classification schemes were explored with different classical machine learning methods. Comparing results with those obtained from the deep learning based methods, such as the proposed cascade learning and ROI-based transfer learning methods, revealed that the performance of the classical machine learning fell behind.

We believe that a large digital dataset with well-represented pneumoconiosis categories will help improve our system significantly. This is especially essential for determining the ILO grade of a chest X-ray positive for pneumoconiosis, rather than simplify distinguishing whether a radiograph is negative or positive for pneumoconiosis. We understand that collecting such a dataset requires time and cooperation among different organizations as the prevalence of pneumoconiosis is assumed to be low in Australia. We have employed CycleGAN to generate synthetic X-ray images for training our machine learning models. The experimental results show that the combination of real and synthetic training data can significantly improve the performance of the machine learning models.

It should be pointed out that there are some limitations with our proposed method. Although our cascade learning based model has achieved high sensitivity and accuracy in detecting pneumoconiosis, it has not been validated on a significant number of pneumoconiosis X-ray images. The system does not yet have functions to quantify the shape and size of opacities according to the ILO Classification System, which is standard information a B-reader is required to

report. While it is encouraging to observe the high sensitivity, specificity and overall classification accuracy in this study, this does not guarantee that our trained machine learning models can be successfully applied to any X-ray images acquired from any X-ray machines. These models have been trained with a mixture of digital radiographs and digitized films from various sources, but the test images used are not representative of all radiographs or X-ray machines. Our machine learning models were trained and tested only with limited number of X-ray images.

References

1. Queensland Government web page, https://www.business.qld.gov.au/industries/mining-energy-water/resources/safety-health/mining/accidents-incidents/mine-dust-lung-diseases, last accessed 2019/05/16.
2. Australian Mining web page: https://www.australianmining.com.au/news/black-lung-case-confirmed-in-nsw/, last accessed 2019/03/18.
3. China Daily, “Miners seek justice for black lung disease”, http://www.chinadaily.com.cn/china/2011-02/16/content_12021599.htm, February 16, 2011, last accessed 2019/03/18.
4. Reuters Health News dated 20 July 2018, https://www.reuters.com/article/us-usa-coal-blacklung/a-tenth-of-u-s-veteran-coal-miners-have-black-lung-disease-niosh-idUSKBN1K92W1, last accessed 2019/03/18.
5. Colinet, J., Rider, J.: Best Practice for Dust Control in Coal Mining. National Institute for Occupational Safety and Health (NIOSH) Publication No. 2010–110 (2010).
6. Roth GA, Abate D, Abate KH, et al. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980–2017: a systematic analysis for the Global Burden of Disease Study 2017. The Lancet 2018; 392(10159):1736-88.
7. Sim M., Glass D., Hoy R., Roberts M., Thompson B., Cohen R.: Review of Respiratory Component of the Coal Mine Workers’ Health Scheme for the Queensland Department of Natural Resources and Mines, Final Report. Monash Centre for Occupational and Environmental Health, Monash University (2016).
8. NIOSH B Reader Certification Program: Looking to the Future. NIOSH Publication No. 2009-140 (2009). https://www.cdc.gov/niosh/docs/2009-140/, last accessed 2018/01/31.
9. The National Institute for Occupational Safety and Health (NIOSH) Chest Radiography: https://wwwn.cdc.gov/niosh-rhd/cwhsp/ReaderList.aspx?formid=InternationalExaminees&lastname=&sortkey=country&format=table&btnSubmit_Intl=Submit, last accessed 2019/05/21.
10. International Labour Office: Guidelines for the use of the ILO International Classification of radiographs of pneumoconiosis. Occupational Safety and Health Series, 22 (2011).
11. Xing J., Huang X., Yang L., Liu Y., Zhang H., Chen W.: Comparison of High-resolution Computerized Tomography with Film-screen Radiography for the Evaluation of Opacity and the Recognition of Coal Workers' Pneumoconiosis. Journal of Occupational Health, 56(4), 301-308 (2014).
12. Sundararajan, R., Xu, H., Annangi, P., Tao, X., Sun, X., and Mao L.: A multiresolution support vector machine based algorithm for pneumoconiosis detection from chest radiographs. In: 2010 IEEE International Symposium on Biomedical Imaging, pp. 1317-1320. IEEE, Rotterdam, The Netherlands (2010).
13. Rajpurkar, P., Irvin J., Zhu K., Yang, B., Mehta H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K., Lungren, M., and Ng, A.: CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. arXiv preprint arXiv:171105225 (2017).
14. Huang, G., Liu, Z., Maaten, L., and Weinberger, K.: Densely Connected Convolutional Networks, pp. 2261-2269. In: IEEE Conference on Computer Vision and Pattern Recognition.Hawaii, USA (2017).
15. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., and Summers, M.: ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and

Localization of Common Thorax Diseases. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3462-3471. Hawaii, USA (2017).
16. Zhu, J., Park, T., Isola, P., and Efros, A.: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. In: IEEE International Conference on Computer Vision (ICCV), pp. 2223-2232. IEEE, Venice, Italy (2017).
17. Kingma, D., and Welling, M.: Auto-encoding variational bayes. arXiv:1312.6114 (2013).
18. Yulia Arzhaeva, Dadong Wang and Deborah Yates, The study protocol for the Coal Services Health and Safety Trust Project No. 20647 - "Development of Automated Diagnostic Tools for Pneumoconiosis Detection from Chest X-Ray Radiographs", Submitted to St Vincent's Hospital Research Office in June 2017 and approved in August 2017.
19. Yulia Arzhaeva, Dadong Wang, CSIRO Human Research Ethics Low Risk Research Project Application Form for the project - "Development of Automated Diagnostic Tools for Pneumoconiosis Detection from Chest X-Ray Radiographs", submitted to CSIRO Human Research Ethics: Low Risk Review Panel and approved in Oct. 2016.
20. The National Institute for Occupational Safety and Health (NIOSH) B Reader Study Syllabus: https://www.cdc.gov/niosh/topics/chestradiography/breader-study-syllabus.html, last accessed 2017/07/04.
21. Japanese Society of Radiological Technology (JSRT) Database: http://db.jsrt.or.jp/eng.php, cited Jan. 22, 2018.
22. 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", American Journal of Roentgenology, 174, 71-74.
23. Van Ginneken B, Stegmann MB, Loog M (2006), "Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database", Medical Image Analysis, 10, 19–40.
24. Wan SitiHalimatulMunirah Wan Ahmad, W Mimi Diyana W Zaki, and Mohammad Faizal Ahmad Fauzi (2015), "Lung segmentation on standard and mobile chest radiographs using oriented Gaussian derivatives filter", Biomed Eng Online, 14, 20. DOI: 10.1186/s12938-015-0014-8.
25. SimpleITK (2017) Image Segmentation and Registration Toolkit, Version 1.0.0: http://www.simpleitk.org, cited Jan. 22, 2018.
26. Scikit-learn (2017) Machine Learning in Python, Version 0.18.2: http://scikit-learn.org/stable/, cited Jan. 22, 2018.
27. Nyul LG, Udupa JK, and Zhang X (2000), "New Variants of a Methods of MRI Scale Standardization", IEEE Transactions on Medical Imaging, 19(2), 143-150.
28. Boser B., Guyon I., and Vapnik V. N.: A Training Algorithm for optimal margin classifiers. In: Proceedings of the Fifth Annual ACM Conference on Computational Learning Theory (COLT 1992), pp. 144 – 152. Pittsburgh, PA, USA (1992).
29. Osuna E., Freund R. and Girosi F., "Training Support Vector Machines: An Application to Face Detection", Proceedings of CVPR'97, Puerto Rico, 1997.
30. Li H, Guan XH, Zan X., "Network intrusion detection based on support vector machine", Journal of Computer Research and Development. 2003; 6:799–807.
31. B. Sch¨olkopf, A. J. Smola, R. Williams, and P. Bartlett, "New support vector algorithms," Neural Computation, vol. 12, pp. 1083–1121, 2000.

32. R. Hadsell, S. Chopra, and Y. Lecun, “Dimensionality reduction by learning an invariant mapping”, in Proc. Computer Vision and Pattern Recognition Conference (CVPR06), IEEE Press, 2006.
33. P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, “Extracting and composing robust features with denoising autoencoders”, in Proceedings of the 25th International Conference on Machine Learning, pages 1096– 1103, 2008.
34. D. P. Kingma and M. Welling, “Auto-encoding variational bayes”, in proceedings of International Conference on Learning Representations, Scottsdale, Arizona, 2-4 May 2013.
35. P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, “Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion”, Journal of Machine Learning Research, 11:3371–3408, 2010.
36. Bernhard Schölkopf and Alexander J Smola, “Support vector machines, regularization, optimization, and beyond”, MIT Press 656 (2002), 657.
37. David MJ Tax and Robert PW Duin, “Support vector data description”, Machine learning 54, 1 (2004), 45–66.
38. Guerbai, Y., Youcef, C., Bilal, H., “The effective use of the one-class SVM classifier for handwritten signature verification based on writer-independent parameters”, Pattern Recognition, Volume 48, Issue 1, January 2015, Pages 103-113.
39. Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou, “Isolation forest”, in Data Mining, ICDM’08. The Eighth IEEE International Conference on Data Mining (ICDM’08), 2008, pp. 413–422.
40. Ho, TK, “Random Decision Forests”, Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16 August 1995. pp. 278–282.
41. Ho, TK, “The Random Subspace Method for Constructing Decision Forests”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 20 (8), 1998, pp 832–844.
42. R Chalapathy, AK Menon, S Chawla, “Anomaly Detection Using One-Class Neural Networks”, Published 2018 in ArXiv: 1802.06360.
43. P. F. Alcantarilla, A. Bartoli, and A. J. Davison, “KAZE features”, In Eur. Conf. on Computer Vision (ECCV), pages 214–227, 2012.
44. P. F. Alcantarilla, J. Nuevo, A. Bartoli, “Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces”, In British Machine Vision Conference (BMVC), 2013.
45. FLANN – Fast Library for Approximate Nearest Neighbours: https://www.cs.ubc.ca/research/flann/, cited Jan. 22, 2018.
46. Li Q, Arimura H, Doi K (2004), “Selective enhancement filters for lung nodules, intracranial aneurysms, and breast microcalcifications”, International Congress Series, 1268: 929-934. https://doi.org/10.1016/j.ics.2004.03.372.
47. Bram van Ginneken, et al., “Multi-scale texture classification from generalized locally orderless images”, Pattern Recognition, vol. 36, pp. 899-911, 2002.
48. Yulia Arzhaeva, et al., “Computer-aided detection of interstitial abnormalities in chest radiographs using a reference standard based on computed tomography”, Medical Physics, vol. 34, no. 12, pp. 4798-4809, 2007.
49. Laurens Hogeweg, et al., “Fusion of Local and Global Detection Systems to Detect Tuberculosis in Chest Radiographs”, MICCAI 2010, Part III, LNCS 6363, pp. 650–657, 2010.

50. Jia Deng, et al. “Imagenet: A large-scale hierarchical image database”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Florida, USA, June 20-25, 2009.

51. Bram van Ginneken, et al., “Automatic detection of abnormalities in chest radiographs using local texture analysis”, IEEE Transactions on Medical Imaging, 21(2), pp. 139-149, 2002.

CONTACT US

t 1300 363 400
+61 3 9545 2176
e csiroenquiries@csiro.au
w www.data61.csiro.au

AT CSIRO WE SHAPE THE FUTURE

We do this by using science and technology to solve real issues. Our research makes a difference to industry, people and the planet.

FOR FURTHER INFORMATION

Dr Dadong Wang
Research Team Leader
t +61 2 93253223
e dadong.wang@csiro.au
w www.data61.csiro.au

Dr Yulia Arzhaeva
Senior Experimental Scientist
t +61 29325 3190
e yulia.Arzhaeva@domain.au
w www.data61.csiro.au

A/Prof. Deborah Yates
Coordinating Principal Investigator
t +61 283822330
e deborahy88@hotmail.com