



Article

# Deep Ensemble Learning for the Automatic Detection of Pneumoconiosis in Coal Worker's Chest X-ray Radiography

Liton Devnath <sup>1,2,\*</sup>, Suhuai Luo <sup>1</sup>, Peter Summons <sup>1</sup> , Dadong Wang <sup>3</sup>, Kamran Shaukat <sup>1,4,\*</sup> , Ibrahim A. Hameed <sup>5,\*</sup> and Fatma S. Alrayes <sup>6</sup>

<sup>1</sup> School of Information and Physical Sciences, The University of Newcastle, Callaghan, NSW 2308, Australia

<sup>2</sup> British Columbia Cancer Research Centre, Vancouver, BC V5Z 1L3, Canada

<sup>3</sup> Quantitative Imaging, CSIRO Data61, Marsfield, NSW 2122, Australia

<sup>4</sup> Department of Data Science, University of the Punjab, Lahore 54890, Pakistan

<sup>5</sup> Department of ICT and Natural Sciences, Norwegian University of Science and Technology, 7491 Trondheim, Norway

<sup>6</sup> Information Systems Department, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University (PNU), P.O. Box 84428, Riyadh 11671, Saudi Arabia

\* Correspondence: liton.devnath@uon.edu.au (L.D.); kamran.shaukat@uon.edu.au (K.S.); ibib@ntnu.no (I.A.H.)

**Abstract:** Globally, coal remains one of the natural resources that provide power to the world. Thousands of people are involved in coal collection, processing, and transportation. Particulate coal dust is produced during these processes, which can crush the lung structure of workers and cause pneumoconiosis. There is no automated system for detecting and monitoring diseases in coal miners, except for specialist radiologists. This paper proposes ensemble learning techniques for detecting pneumoconiosis disease in chest X-ray radiographs (CXRs) using multiple deep learning models. Three ensemble learning techniques (simple averaging, multi-weighted averaging, and majority voting (MVOT)) were proposed to investigate performances using randomised cross-folds and leave-one-out cross-validations datasets. Five statistical measurements were used to compare the outcomes of the three investigations on the proposed integrated approach with state-of-the-art approaches from the literature for the same dataset. In the second investigation, the statistical combination was marginally enhanced in the ensemble of multi-weighted averaging on a robust model, CheXNet. However, in the third investigation, the same model elevated accuracies from 87.80 to 90.2%. The investigated results helped us identify a robust deep learning model and ensemble framework that outperformed others, achieving an accuracy of 91.50% in the automated detection of pneumoconiosis.

**Keywords:** ensemble learning; pneumoconiosis; transfer learning; deep learning; chest X-ray radiographs; cross validation; weighted-averaging; majority voting; LOOCV; CheXNet



**Citation:** Devnath, L.; Luo, S.; Summons, P.; Wang, D.; Shaukat, K.; Hameed, I.A.; Alrayes, F.S. Deep Ensemble Learning for the Automatic Detection of Pneumoconiosis in Coal Worker's Chest X-ray Radiography. *J. Clin. Med.* **2022**, *11*, 5342. <https://doi.org/10.3390/jcm11185342>

Academic Editor: Filippo Pesapane

Received: 19 July 2022

Accepted: 7 September 2022

Published: 12 September 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Deep learning models are susceptible to noise in training data, as they learn by using stochastic gradient functions. This causes variance errors and may cause overfitting, resulting in low generalisations for validating data. A machine learning technique known as ensemble learning reduces predictive variance by combining the predictions of integrated models. Ensembles are often more accurate than individual classifiers that produce them [1–5].

On the other hand, a deep convolutional neural networks (CNNs) drive process is a difficult optimising process that often does not converge. As a result, CNN's latest drive weights may not show a consistent or optimal performance as the final model weights. To overcome this problem, the average performance of training weights is calculated as many points in the training cycle [6–8]. In general, it could be called the average weight prediction based on the method developed by Polyak-Ruppert [9,10].

Additionally, every CNN is very sensitive to the volume of training data. The model will learn better if you have high-volume data. A special case of cross-validation is called leave-one-out cross-validation (LOOCV), and it is used to evaluate the efficiency of machine learning models with a small dataset. This is a lengthy and costly process, even though it provides a reliable and impartial estimate of model performance. While very simple in application, there are some limitations in using, as there is no need for its application if a large dataset or mathematically costly method is used. During the application of the LOOCV process, each machine learning model is adjusted at a higher number of times, representing a more robust assessment since each data can participate as the entire test dataset [11,12].

In recent years, deep transfer learning with an ensemble of multiple CNNs has been widely used in medical-image processing [13–17]. The trained deep ensemble learning represents a single hypothesis. Empirically, ensembles yield better results when significant diversity among the models, even on a small dataset. Therefore, many ensemble methods seek to promote diversity among the combined models. An ensemble indicates different techniques, including simple averaging, weighted-averaging, majority voting (MVOT), bagging, boosting, CNN blocks, randomizing, and stacking using multi-model predictions on the same dataset [18–21].

This paper proposed simple averaging, weighted-averaging, and MVOT techniques to detect pneumoconiosis in coal workers' chest X-ray radiographs (CXRs). The summary of our list of contributions is as follows:

- I. We have used databases of posterior-anterior (PA) CXRs collected from various hospitals by the Commonwealth Scientific and Industrial Research Organisation (CSIRO), Sydney, Australia. To overcome the problems associated with small datasets, we assessed proposed ensemble techniques, simple averaging, weighted-averaging, and MVOT using randomised cross-fold-validation (RCFV) and leave-one-out cross-validations (LOOCV) of the original dataset independently.
- II. In all techniques, transfer learning has been implemented using multiple CNN are namely CheXNet [22], DenseNet-121 [23], Inception-V3 [24], Xception [25], and ResNet50 [26]. We proposed ensemble techniques in three investigations: investigation-1 uses simple averaging on RCFV data, investigation-2 uses weighted averaging on RCFV data, and investigation-3 uses MVOT on LOOCV data.
- III. Finally, we compared the investigation's outcomes using five formulas of statistical measurements [27], sensitivity, specificity, accuracy, precision, and F1-Score, with state-of-the-art approaches from the literature for the same dataset and highlighted the efficient CNN model in our dataset.

The following Figure 1 depicts the overall contributions, providing an improved understanding of what we have performed in this study. Section 2 presents background studies and findings for pneumoconiosis classification on the same dataset using various classical, traditional machine, and deep learning methods. The orientation of the dataset and the detailed methodologies within each investigation are presented separately in Section 3. Section 4 provides the outcomes of investigation-1, investigation-2, and investigation-3. Section 5 summarised the outcomes of the investigation and compared them with state-of-the-art approaches from the background study for the same dataset. The assumptions and limitations are also highlighted there. Finally, Section 6 provides the conclusion of this research study.



**Figure 1.** Summary of proposed methodologies in the study of ensemble investigations.

## 2. Background Study

The abnormality on a chest X-ray of the lung is signified by the increase or decrease in density areas. The chest X-ray lung abnormalities with increased density are also known as pulmonary opacities. Pulmonary opacities have three major patterns: consolidation, interstitial, and atelectasis. Among these, the interstitial patterns of pulmonary opacities are mainly responsible for pneumoconiosis disease [28]. According to the International Labour Organization’s (ILO) classification, two abnormalities are observed for all types of pneumoconiosis—parenchymal and pleural. Parenchymal abnormalities are indicated by small opacity shape (round or irregular) and size ( $1.5\text{ mm} < \text{diameter (round)} < 10\text{ mm}$  and  $1.5\text{ mm} < \text{widths (irregular)} < 10\text{ mm}$ ) and large opacities of a round shape and size less than or equal to 50 mm. Pleural abnormalities are mainly indicated by angle obliteration and the diffusion of thickness in the CXR’s wall [29].

There is no national approach to health screening of coal miners in Australia. In NSW, a chest X-ray is recommended every six years for mine-site workers but it is not mandatory.

Medical screening has also failed to detect this potentially fatal disease [30]. For these reasons, it is desirable to develop an established computer-based automatic system further to provide the quantitative evaluation of pneumoconiosis and serve as an initial screening process and a second opinion for medical doctors.

Past research on the automatic classification of pneumoconiosis classical, traditional machine, and deep learning methods were used. The texture features were mostly classified in classical methods using computer- and ILO-based standard classification [31–41]. The profusion of small round opacities and ILO extent properties indicated normal and abnormal lungs. The backpropagation neural networks have been applied to find the shape and size of round opacities from the region of interest (ROI) portions of an image [42–45]. X-ray abnormalities were categorised and compared with the results of the standard ILO measurement of the size and shape of the round opacities.

In traditional machine learning, different methods for handcrafted feature extraction, or selection were used. The handcrafted features, such as texture features [46,47], from the left–right lung zones [48–51] were extracted. After the selection of important features, they were input into different machine learning classifiers, such as support vector machine (SVM) [49,52–60], decision trees (DT) [55,56], random trees (RT) [57–60], artificial neural networks (ANNs) [61–63], K-nearest neighbours (KNN) [64], self-organizing map (SOM) [64], backpropagation (BP), radial basis function (RBF) neural networks (NN) [57–60,64,65], and ensemble classifier [49,52,56].

In recent years, deep learning approaches have achieved state-of-the-art results due to their high dimensional feature representation of data [66,67]. Many deep convolutional neural networks performed better than humans, especially in medical image processing [68]. Such examples include identifying indicators for cancer in blood [69] and skin [70,71], malaria in blood cell [72], tuberculosis (TB) from chest X-rays [14,16,73], and more specifically pneumoconiosis in chest X-rays [27,74–80].

We have conducted different classical, traditional, and deep learning approaches in our previous published works on the same dataset used in this study. We used the ILO Standard Classification System in classical approaches, and the performance is presented in Table 1.

**Table 1.** Summary of all classical, traditional, and deep learning approaches previously performed on the same dataset.

Year	Ref No.	Dataset	Classification Approaches	Evaluation Performance		
				Accuracy	Specificity	Recall
2019	[78]	Same dataset that was used in this paper	Classical method, ILO standard	83.00%	81.70%	84.60%
			SVM = 73.17%	92.31%	73.30%	
			MLP = 71.11%	72.00%	70.00%	
			NN = 83.00%	85.00%	82.00%	
2019	[78]		Traditional machine learning classifiers	Isolation Forest = 73.30%	92.31%	73.17%
				KNN = 69.30%	-	-
				Random Forest = 70.80%	-	-
				Ridge = 76.90%	87.00%	63.00%
2019	[76]		CNN-without transfer learning	DenseNet = 80.49%	66.67%	88.46%
				VGG16 = 82.93%	80.00%	84.62%
				VGG19 = 80.49%	80.00%	80.77%
				ResNet = 85.37%	80.00%	88.46%
2020	[77]	153 CXR including 71 pneumoconiosis	Deep CNN-transfer learning	InceptionV3 = 87.80%	86.67%	88.46%
				Xception = 85.37%	93.33%	80.77%
				DenseNet = 82.93%	80.00%	84.62%
				CheXNet = 85.37%	93.33%	80.77%
2019	[75]		Cascaded Learning	90.24%	88.46%	93.33%

We first extracted handcrafted features using different statistical image analysis methods in traditional machine learning approaches. Then, we input these features into different machine learning classifiers, such as support vector machine (SVM), MLP, NN, K-nearest neighbours (KNN), isolation forest, random forest, and ridge [78]. We show these classifier results in Table 1.

In deep learning approaches, first, we implemented, with and without transfer learning, convolutional neural networks (CNN) to detect pneumoconiosis. Deep transfer learning was implemented using seven pre-trained CNNs, VGG16 [81], VGG19, Inception [24], Xception [25], ResNet50 [26], DenseNet-121 [23], and CheXNet [22]. Then, we performed a performance comparison between them. The comparison was examined using different effects of dropout rates and different augmentation methods used in DL models, with and without transfer learning, to detect pneumoconiosis. We developed a cascade learning model, which outperforms others and achieved an overall classification accuracy of 90.24%, a specificity of 88.46%, and a sensitivity of 93.33% for detecting pneumoconiosis using generated synthesised images from real segmented CXR databases. We have also summarised deep CNNs results in Table 1. The previous studies showed that the deep transfer learning performance of Inception-V3, Xception, ResNet50, DenseNet, and CheXNet was satisfactory compared to classical and traditional approaches.

### 3. Datasets and Methods

The first part of this section discusses our dataset and how it was processed using cross-validation to perform ensemble techniques. In contrast, the rest of the section describes the techniques used in three investigations.

#### 3.1. Materials

Out of a collaboration between the University of Newcastle and the Commonwealth Scientific and Industrial Research Organisation (CSIRO) data61 Sydney NSW, Australia, chest X-ray image datasets were built with associated diagnostic labels for this study. CSIRO data61 collected the data from Coal Services Health NSW, St Vincent's Hospital, Sydney, and Wesley Medical Imaging, Queensland. The publicly available NIOSH teaching chest X-ray dataset and ILO Standard Radiographs (International Labour Organization, (ILO) Genève, Switzerland) were also used to develop parts of the small dataset DL model. All radiographs used in this study are posterior-anterior (PA) radiographs. Seventy-one PA chest radiographs with small parenchymal opacities consistent with pneumoconiosis and 82 PA chest X-rays belonging to normal individuals were used. All data were collected from coal mine workers, including males and females. We conducted ensemble learning using randomised cross-fold-validation and leave-one-out cross-validation. The details are in the following subsections:

##### 3.1.1. Randomised Cross-Fold-Validation

To maintain the balance of training data, 112 X-rays (56 normal and 56 pneumoconiosis) were used for training and 41 X-rays (26 normal and 15 pneumoconiosis) were used for testing. Twenty-five percent of training data were kept as a validation dataset for selecting the best model weights based on validation performance. We continued the randomised selection three times and then organised our total dataset into three different folds, namely, as randomised cross-fold-validation (RCFV) dataset 1, dataset 2, and dataset 3, as shown in Figure 2. Therefore, we defined this cross-validation simply as RCFV.

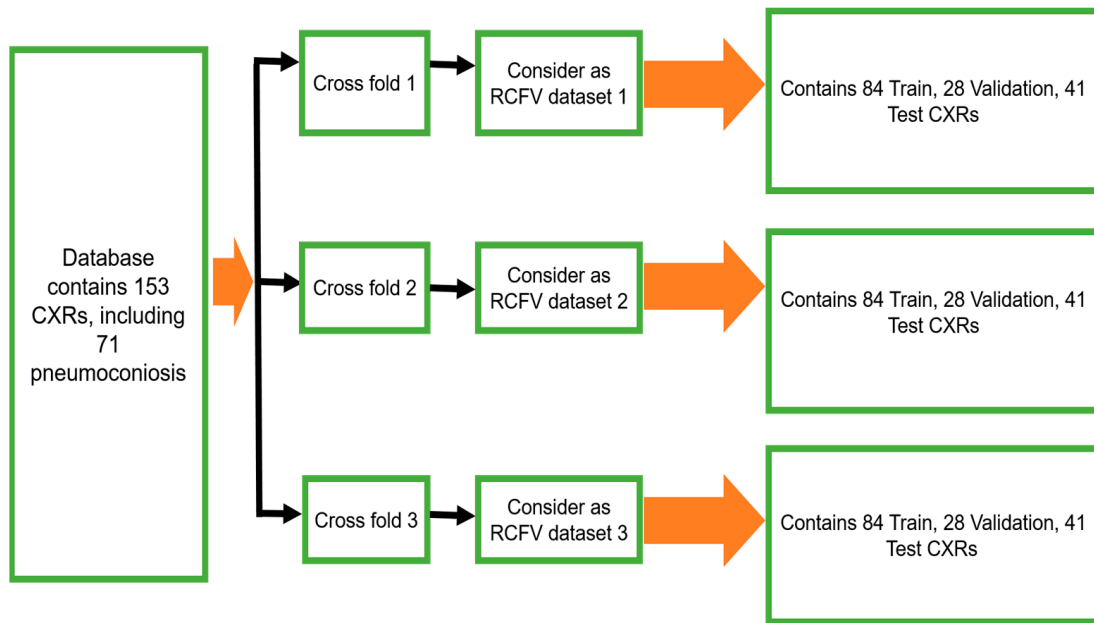


Figure 2. Three RCFVs of our proposed dataset.

3.1.2. Leave-One-Out Cross-Validation

We proposed a case of cross-validation (LOOCV), which is used to assess the effectiveness of machine learning models with the same dataset. We organised our dataset into two groups, dataset A and dataset B, clearly mentioned in Figure 3. Dataset A contained 71 pairs of images, including an equal number of normal and abnormal (pneumoconiosis) CXRs. Therefore, the remaining 11 normal images were in dataset B. As a result, no correlation exists between the pairs of images.

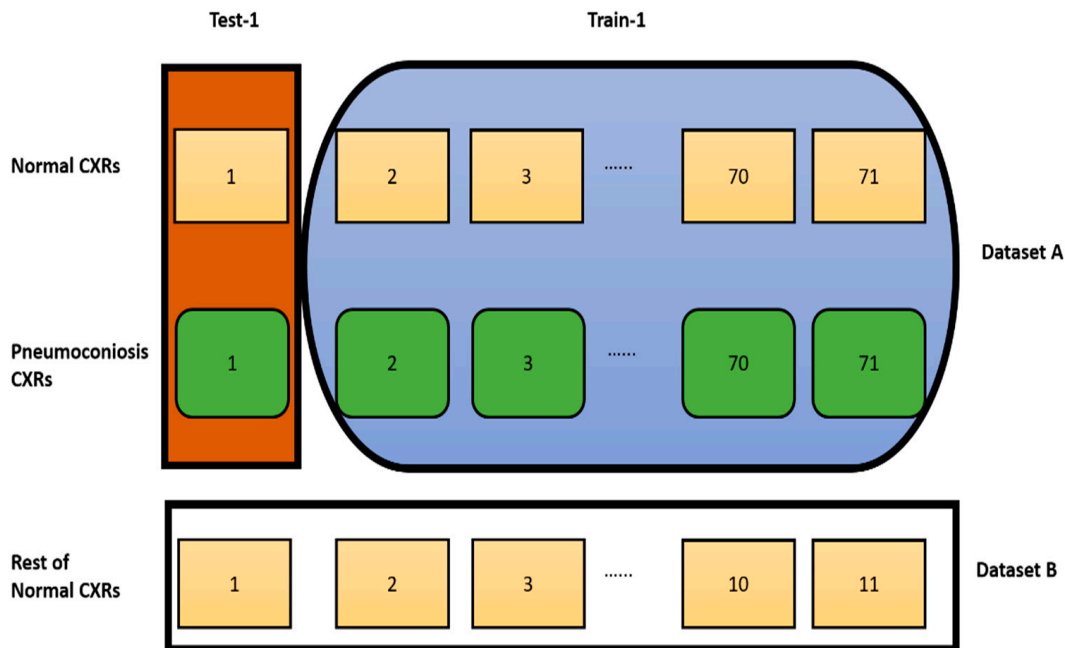


Figure 3. Data orientation for LOOCV implementation.

3.2. Methods

The proposed ensemble techniques, simple averaging, weighted averaging, and MVOT, were independently conducted using RCFV and LOOCV datasets. In all tech-



niques, transfer learning was analysed by the same CNNs, namely CheXNet, DenseNet-121, Inception-V3, Xception, and ResNet50. We organized our proposed method into three investigations, as stated below.

### 3.2.1. Investigation-1: An Ensemble Learning Using Simple Averaging through RCFV Datasets

The deep learning model shows each test element's probability value within the range [0, 1] during the forecast. Those fractional probabilities are converted to predict class labels using a threshold value condition. An ensemble is mapped using several CNNs model prediction probabilities as a combined decision instead of individually. Therefore, each value of the testing data was predicted by multiple models at once. After that, their average predictive probability interval between [0, 1] indicates the ensemble's performance.

In this investigation, we implemented deep transfer learning throughout the ensemble using simple averaging of the probability of detection of pneumoconiosis predicted by five CNN models: CheXNet, DenseNet, Inception-V3, Xception, and ResNet50. Afterwards, we calculated the average prediction probabilities on the same RCFV testing datasets 1 to 3, as demonstrated in Figure 2.

CNNs employ a stochastic learning algorithm to optimise training randomly. The optimisation is based on selecting the loss function while the model has been designed. The purpose of the loss function is to determine whether the model is operating properly or incorrectly. The cost function within the CNNs determines the difference in losses between true and predicted values. We applied the regularisation technique to reduce the complexity of a neural network model during training and, thus, prevented overfitting. There are very popular and efficient regularisation techniques called L2. The regularisation term is weighted by the scalar lambda divided by 2 m and added to the regular loss function chosen for the current task. This leads to a new expression for the loss function, as shown in the following Equation (1):

$$\text{Cost function} = \text{loss}(\text{binary\_cross\_entropy}) + \frac{\lambda}{2m} \sum \|w^2\| \quad (1)$$

where  $\lambda$  denotes the regularization parameter, and its value may optimise the learning rate for improved predictions. L2 regularisation is also known as the weight decay as it forces the weights to decay towards zero (but not exactly zero).

After taking the output of each of the five models, one GlobalAveragePooling2D layer was added. Three dense layers, with all their output nodes, were connected with all nodes of the next layer. Global Average Pooling is a transaction that computes the average performance of each entity map in the preceding layer. This relatively simple operation helps convert the data into a one-dimensional vector and avoids the overflow of features. There are no trainable parameters, similarly to the Max pooling operation.

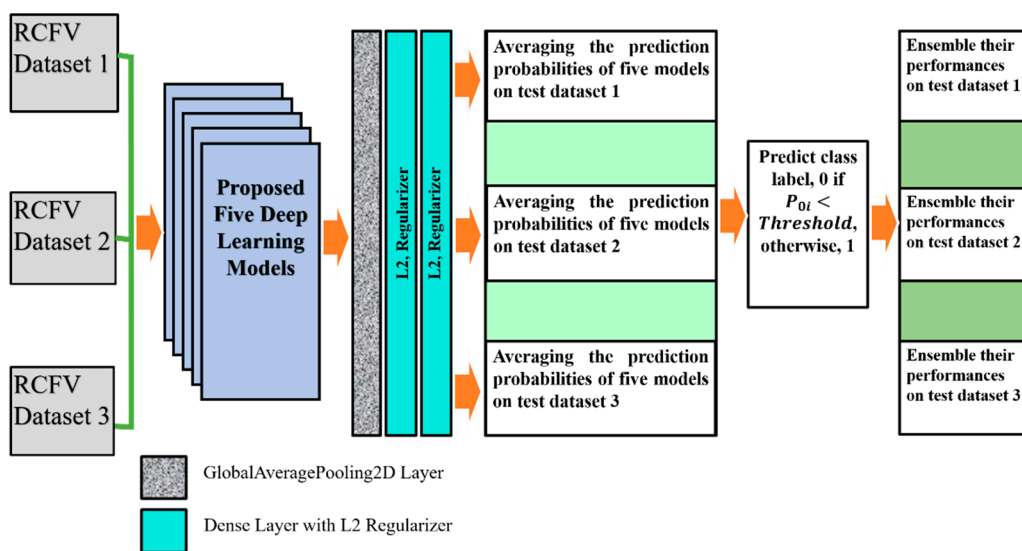
For L2 (0.001), two regularisers were used with the first-two dense layers for better optimisation with the proposed models. The last layer of the classifier used a sigmoid activation function and output probability scores for each class—normal and pneumoconiosis (see Figure 4). We used  $512 \times 512$  X-ray input forms for each proposed CNNs architecture, where the output of the prediction probability value ranged between [0, 1]. The regular loss-function, binary cross-entropy with an Adam optimiser of the learning rate, 0.0001, was also used during training.

We trained each DL model up to 50 epochs and used the last weights to find the prediction probability of normal and abnormal CXRs. For instance, in RCFV dataset-1, we applied five models independently and then calculated their prediction probabilities separately. Next, we calculated the average of their probability values for each unique test image using mathematical Equation (2). If the average value  $P_{0i} < \text{Threshold}(0.5)$ , then

its predicted label changes to 0; otherwise, it is 1, where  $i = 1$  to 26 for normal images and  $i = 1$  to 15 for pneumoconiosis images.

$$P_{0i} = \frac{Model1_{0i} + Model2_{0i} + Model3_{0i} + Model4_{0i} + Model5_{0i}}{\text{total number of models}} \quad (2)$$

The ensemble performances of five models, CheXNet, DenseNet, Inception-V3, Xception, and ResNet50, were computed using confusion matrix values, true positives, false negatives, true negatives, and false positives. The ensemble performance for RCFV datasets 2 and 3 was calculated according to the same process used for dataset 1. The details of the proposed workflow are demonstrated in Figure 4. The last three columns illustrate the direction of the average probability forecasts, the forecast labels, and the ensemble performance of the five models across three different cross-datasets.



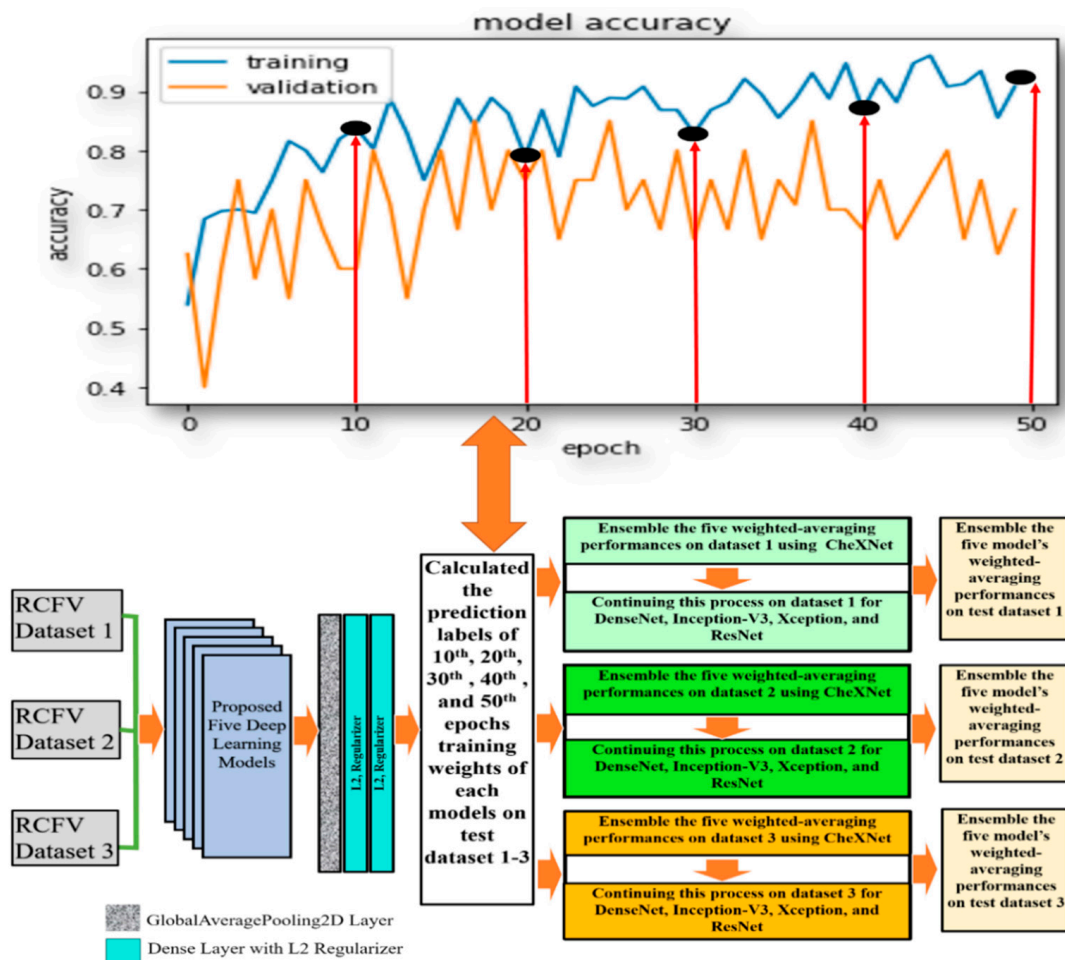
**Figure 4.** An ensemble learning based on simple averaging of probability prediction values using multiple DL models in three RCFV datasets.

### 3.2.2. Investigation-2: An Ensemble Learning Using Weighted Averaging through RCFV Datasets

With the method used in the previous investigation, we investigated multi-model ensemble learning using the latest drive weights of each model in detecting pneumoconiosis diseases from CXR. We replicate the investigation in this section using the same five models, CheXNet, DenseNet, Inception-V3, Xception, and ResNet50, used in investigation-1. To find the optimal solution for pneumoconiosis detection, we carried out ensemble learning using the combination of the weighted average and majority voting techniques. Here, we focused on its different training epochs in calculating the weighted average ensemble for a single model. We kept the same training process and dataset, as described in investigation-1. In calculating a weighted average on a single model, we used the specified weights from the epochs (10th, 20th, 30th, 40th, and 50th set) for each proposed model, as defined in the central white box in Figure 5.

For instance, in dataset 1, we trained the CheXNet model independently on training data and then computed the five sets of prediction labels of its 10th, 20th, 30th, 40th, and 50th epochs' weights with test data. The weighted average ensemble prediction labels of the CheXNet were found using the majority voting (MVOT) decision on these five sets of predictions. As a result, if, and only if, the weight of the majority says that a BL image is BL, then the ensemble decision is BL; otherwise, it is normal.





**Figure 5.** An ensemble using the average prediction probabilities of the combined five DL models on three different RCFV datasets.

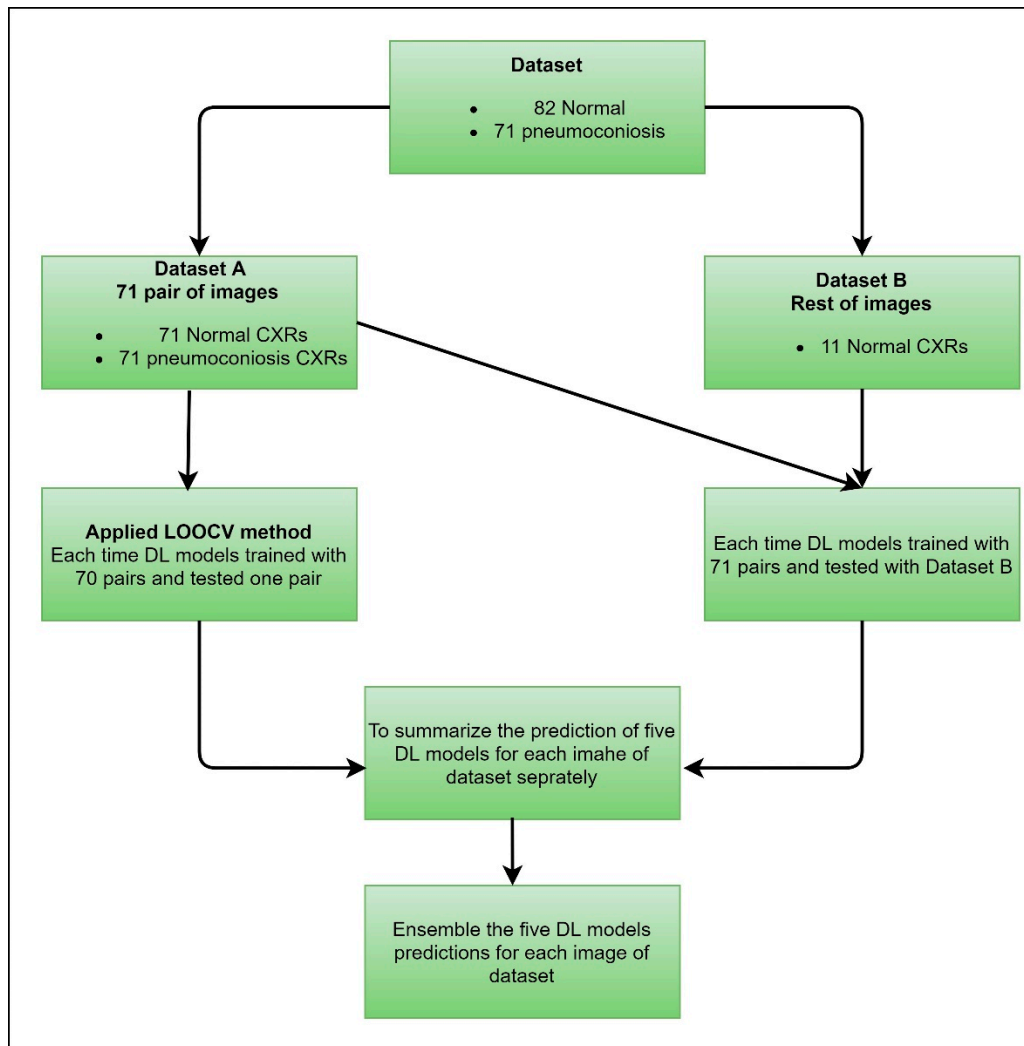
Likewise, we continued this process for dataset 1 for DenseNet-121, Inception-V3, Xception, and ResNet50 models, as described in the second last column in Figure 5. Finally, every single model’s weighted average ensemble return was used to calculate the multi-model ensemble for dataset-1. To accomplish this, MVOT was also applied to the five independent sets of five weighted average prediction labels in the models, as described in the last column of Figure 5.

Similarly, we conducted this process for the testing datasets 2 and 3 and compared weighted-averaging ensemble performances of a single and integrated model using true positive, false negative, true negative, and false positive values from the predicted confusion matrix.

### 3.2.3. Investigation-3: An Ensemble Learning Using MVOT through the LOOCV Dataset

In this investigation, we implemented LOOCV to select a robust DL model from CheXNet, DenseNet, InceptionV3, Xception, and ResNet50 by using our organised dataset, as discussed in Section 3.1.2, representing the best competence for detecting detection pneumoconiosis from CXRs. Figure 6 shows how we handled training and testing using DL applications for each cross-data application. For dataset A, every DL model has trained on 70 pairs of images and tested one pair. We continued the process 71 times automatically. We trained the same model with dataset A for dataset B and then tested the performances on dataset B. Next, we independently calculated each DL model performance for each image using a combination of datasets A and B. Finally, each model’s predictions ensemble return is used to calculate the multi-model ensemble for all data in LOOCV using a simple MVOT

technique. Therefore, if the majority of models predict as “normal”, then its ensemble prediction is defined as a “normal” or, otherwise, “abnormal” lung.



**Figure 6.** Applying the LOOCV method for the detection of pneumoconiosis using a deep learning algorithm.

Finally, we compared the MVOT-based ensemble performances of a single and integrated model using true positive, false negative, true negative, and false positive values from the predicted confusion matrix.

#### 4. Results

This section provides a detailed outcome of the three methodological investigations conducted sequentially.

##### 4.1. In Investigation-1

We independently applied five deep CNNs models (CheXNet, DenseNet, Inception-V3, Xception, and ResNet50) on RCFV datasets 1–3. The regularisation technique was also implemented for an improved optimisation of CNN learning. We used 84 images (equal class of normal and pneumoconiosis) for training, 28 images (equal class of normal and pneumoconiosis) for validation, and 41 images (26 normal and 15 pneumoconiosis) for testing each model.

We calculated the testing probability of a single image within RCFV datasets 1 to 3. We then converted each fractional value into class label 0 or 1 based on the Threshold (0.5), as shown in Figure 4. Table 2 demonstrates the performance based on the prediction probability of five DL models separately on three different datasets. Each model’s performance was evaluated with the metrics values, including sensitivity, specificity, accuracy, precision, and F1-Score.

**Table 2.** The performance is based on the prediction probability of five CNNs models separately on three RCFV datasets.

RCFV Dataset	Model	Sensitivity (%)	Specificity (%)	Accuracy (%)	Precision (%)	F1-Score (%)
1	<b>CheXNet</b>	<b>83.33</b>	<b>90.91</b>	<b>85.37</b>	<b>96.15</b>	<b>89.29</b>
	DenseNet	84.00	68.75	78.05	80.77	82.35
	InceptionV3	76.47	100.00	80.49	100.00	86.67
	Xception	71.88	66.67	70.73	88.46	79.31
	Resnet50	71.43	53.85	65.85	76.92	74.50
2	<b>CheXNet</b>	<b>83.33</b>	<b>90.91</b>	<b>85.37</b>	<b>96.15</b>	<b>89.29</b>
	DenseNet	78.57	69.23	75.61	84.62	81.48
	InceptionV3	82.14	76.92	80.49	88.46	85.19
	Xception	90.91	68.42	80.49	76.92	83.33
	Resnet50	100.00	60.00	75.61	61.54	76.19
3	<b>CheXNet</b>	<b>80.00</b>	<b>81.82</b>	<b>80.49</b>	<b>92.31</b>	<b>85.71</b>
	DenseNet	72.00	50.00	63.41	69.23	70.59
	InceptionV3	73.53	85.71	75.61	96.15	83.33
	Xception	85.71	60.00	73.17	69.23	76.60
	Resnet50	75.00	77.78	75.61	92.31	82.76

In Tables 3–5, we demonstrate the five-model prediction probability on the specified columns. Afterwards, we calculated the average prediction value using Equation (2) for each testing image of datasets 1–3. The rightmost two columns represent each image’s predict and true labels separately. The prediction label was calculated based on the average prediction rate of five models using respective test datasets. The true label column indicates that the first 26 and last 15 images were normal and pneumoconiosis classes.

**Table 3.** Average testing probabilities of five models on RCFV dataset 1.

Testing Img No.	CheXNet	DenseNet	Inception V3	Xception	ResNet50	Average of Five Models
1	0.932637	0.040125	0.450016	0.025675	0.002168	0.290125
2	0.105605	0.007986	0.028912	0.002934	0.000006	0.029089
3	0.039132	0.257459	0.089236	0.032493	0.000001	0.083665
4	0.052019	0.325931	0.162149	0.227888	0.000924	0.153783
5	0.036418	0.462622	0.294643	0.727933	0.999364	0.504196
6	0.124626	0.002598	0.092452	0.001123	0.000331	0.044227
7	0.122279	0.004001	0.055866	0.000066	0.000010	0.036445
8	0.238593	0.001707	0.064059	0.014911	0.000269	0.063908
9	0.178124	0.976175	0.122117	0.839697	0.999989	0.623221
10–41	0.336297	0.459394	0.277296	0.257559	0.379978	0.342105

**Table 4.** Average testing probabilities of five models on RCFV dataset 2.

Testing Img No.	CheXNet	DenseNet	InceptionV3	Xception	ResNet50	Average of Five Models
1	0.272361	0.001067	0.000956	0.028802	0.001717	0.060981
2	0.579127	0.877206	0.837060	0.994562	0.999967	0.857585
3	0.360972	0.002431	0.001269	0.007587	0.632691	0.200990
4	0.377410	0.000500	0.014508	0.529541	0.847036	0.353799
5	0.418817	0.008354	0.004357	0.000794	0.303923	0.14725
6	0.470152	0.000417	0.002373	0.001013	0.209452	0.136682
7	0.113039	0.005697	0.016339	0.984138	0.387843	0.301412
8	0.364287	0.007771	0.000192	0.010377	0.007184	0.077963
9	0.223685	0.003317	0.596456	0.956602	0.977765	0.551566
10–41	0.417755	0.395281	0.362951	0.464778	0.661663	0.460483

**Table 5.** Average testing probabilities of five models on RCFV dataset 3.

Testing Img No.	CheXNet	DenseNet	InceptionV3	Xception	ResNet50	Average of Five Models
1	0.669168	0.099058	0.191127	0.981978	0.161130	0.420493
2	0.387482	0.002331	0.011922	0.001326	0.000510	0.080715
3	0.334752	0.218163	0.033087	0.002177	0.000000	0.117636
4	0.425194	0.985000	0.265910	0.996532	0.716426	0.677813
5	0.192840	0.998790	0.241946	0.997667	0.996475	0.685544
6	0.455117	0.805958	0.013506	0.000174	0.000158	0.254983
7	0.431945	0.279125	0.009881	0.000019	0.000255	0.144245
8	0.530023	0.673950	0.136910	0.250284	0.005652	0.319364
9	0.464333	0.000178	0.001372	0.000228	0.000017	0.093226
10–41	0.431321	0.376997	0.244273	0.524577	0.228047	0.361043

We calculated the confusion metric values, true positive (TP), true negative (TN), false positive (FP), and false negative (FN) for every dataset by counting the predict and true labels from Table 3 to Table 5.

We demonstrated ensemble learning performances on five models’ prediction probabilities using eight evaluation metrics in Table 6, in which the ensemble was learning of the model’s prediction probability represented maximum values of sensitivity, specificity, accuracy, precision, and F1-Score, with a sensitivity of 88.00%, a specificity of 75.00%, an accuracy of 82.93%, a precision of 84.62%, and an F1-score of 86.27% for dataset 2, which are lower values than the individual performances of the CheXNet model without ensemble learning. The performance of ensemble learning using the model’s prediction probability did not improve pneumoconiosis’ detection accuracy.

**Table 6.** An ensemble using the average prediction probabilities of combining five DL models on three RCFV datasets.

RCFV Dataset	Ensemble of Models	Sensitivity (%)	Specificity (%)	Accuracy (%)	Precision (%)	F1-Score (%)
1	CheXNet, DenseNet,	69.70	62.50	68.29	88.46	77.97
2	InceptionV3,	88.00	75.00	82.93	84.62	86.27
3	Xception, Resnet50	82.14	76.92	80.49	88.46	85.19

#### 4.2. In Investigation-2

The deep learning models, CheXNet, DenseNet, Inception-V3, Xception, and ResNet50, have been used to calculate the prediction labels of their trained weights of the 10th, 20th,

30th, 40th, and 50th epochs' iteration, as demonstrated in Figure 5. We used the same training, validation, and testing datasets previously in investigation-1 and evaluated five trained weights and their ensemble performances using the same metrics' values. All assessments of a single weight and its ensemble were presented separately for each model in the three RCFV cross-fold datasets.

In Table 7, we have represented the CheXNet performances of specified weights with ensemble learning on three RCFV datasets, 1–3. The five trained weights have shown different sensitivity, specificity, accuracy, precision, and F1-Score measurements within three datasets. Ensemble learning shows that the CheXNet achieved a sensitivity of 86.21%, a specificity of 91.67%, an accuracy of 87.80%, a precision of 96.15%, and an F1-score of 90.91% for dataset 1. For dataset 2, CheXNet achieved a sensitivity of 83.87%, a specificity of 100.00%, an accuracy of 87.80%, a precision of 100.00%, and an F1-score of 91.23%. It is also shown that sensitivity of 78.13%, a specificity of 88.89%, an accuracy of 80.49%, a precision of 96.15%, and an F1-score of 86.21% were achieved for dataset 3.

**Table 7.** The CheXNet performances of specified weights with ensemble learning.

RCFV Dataset	CheXNet Trained Weights	Sensitivity (%)	Specificity (%)	Accuracy (%)	Precision (%)	F1-Score (%)
1	10-epoch	88.89	85.71	87.8	92.31	90.57
	20-epoch	80.65	90.00	82.93	96.15	87.72
	30-epoch	83.33	90.91	85.37	96.15	89.29
	40-epoch	83.33	90.91	85.37	96.15	89.29
	50-epoch	83.33	90.91	85.37	96.15	89.29
	<b>Ensemble learning</b>	<b>86.21</b>	<b>91.67</b>	<b>87.80</b>	<b>96.15</b>	<b>90.91</b>
2	10-epoch	80.77	66.67	75.61	80.77	80.77
	20-epoch	83.33	90.91	85.37	96.15	89.29
	30-epoch	80.00	81.82	80.49	92.31	85.71
	40-epoch	74.29	100.00	78.05	100.00	85.25
	50-epoch	72.22	100.00	75.61	100.00	83.87
	<b>Ensemble learning</b>	<b>83.87</b>	<b>100.00</b>	<b>87.80</b>	<b>100.00</b>	<b>91.23</b>
3	10-epoch	80.00	81.82	80.49	92.31	85.71
	20-epoch	73.53	85.71	75.61	96.15	83.33
	30-epoch	71.43	83.33	73.17	96.15	81.97
	40-epoch	71.43	83.33	73.17	96.15	81.97
	50-epoch	67.57	75.00	68.29	96.15	79.37
	<b>Ensemble learning</b>	<b>78.13</b>	<b>88.89</b>	<b>80.49</b>	<b>96.15</b>	<b>86.21</b>

In Table 8, we have represented the DenseNet performances of specified weights with ensemble learning on three RCFV datasets, 1–3. The five trained weights have shown different sensitivity, specificity, accuracy, precision, and F1-Score measurements within the three datasets. Ensemble learning has demonstrated that the DenseNet achieved a sensitivity of 80.77%, a specificity of 66.67%, an accuracy of 75.61%, a precision of 80.77%, and an F1-score of 80.77% for dataset 1. For dataset 2, DenseNet achieved a sensitivity of 79.31%, a specificity of 75.00%, an accuracy of 78.05%, a precision of 88.46%, and an F1-score of 83.64%. It has also shown a sensitivity of 78.57%, a specificity of 69.23%, an accuracy of 75.61%, a precision of 84.62%, and an F1-score of 81.48% achieved for dataset 3.

**Table 8.** DenseNet performances of specified weights with ensemble learning.

RCFV Dataset	DenseNet Trained Weights	Sensitivity (%)	Specificity (%)	Accuracy (%)	Precision (%)	F1-Score (%)
1	10-epoch	80.77	66.67	75.61	80.77	80.77
	20-epoch	72.22	100.00	75.61	100.00	83.87
	30-epoch	76.67	72.73	75.61	88.46	82.14
	40-epoch	85.00	57.14	70.73	65.38	73.91
	50-epoch	84.00	68.75	78.05	80.77	82.35
	<b>Ensemble learning</b>	<b>80.77</b>	<b>66.67</b>	<b>75.61</b>	<b>80.77</b>	<b>80.77</b>
2	10-epoch	86.96	66.67	78.05	76.92	81.63
	20-epoch	95.45	73.68	85.37	80.77	87.50
	30-epoch	76.47	100.00	80.49	100.00	86.67
	40-epoch	80.65	90.00	82.93	96.15	87.72
	50-epoch	78.57	69.23	75.61	84.62	81.48
	<b>Ensemble learning</b>	<b>79.31</b>	<b>75.00</b>	<b>78.05</b>	<b>88.46</b>	<b>83.64</b>
3	10-epoch	75.00	77.78	75.61	92.31	82.76
	20-epoch	72.22	100.00	75.61	100.00	83.87
	30-epoch	80.00	62.50	73.17	76.92	78.43
	40-epoch	72.00	50.00	63.41	69.23	70.59
	50-epoch	80.00	62.50	73.17	76.92	78.43
	<b>Ensemble learning</b>	<b>78.57</b>	<b>69.23</b>	<b>75.61</b>	<b>84.62</b>	<b>81.48</b>

In Table 9, we demonstrated the InceptionV3 performances of specified weights with ensemble learning on three RCFV datasets, 1–3. The five trained weights have shown different sensitivity, specificity, accuracy, precision, and F1-Score measurements within the three datasets. Ensemble learning has shown that the InceptionV3 achieved a sensitivity of 85.71%, a specificity of 84.62%, an accuracy of 85.37%, a precision of 92.31%, and an F1-score of 88.89% for dataset 1. For dataset 2, the InceptionV3 achieved a sensitivity of 88.89%, a specificity of 85.71%, an accuracy of 87.80%, a precision of 92.31%, and an F1-score of 90.57%. It has also shown that sensitivity of 74.29%, a specificity of 100.00%, an accuracy of 78.05%, a precision of 100.00%, and an F1-score of 85.25% were achieved for dataset 3.

In Table 10, we demonstrated the Xception performances of specified weights with ensemble learning on three RCFV datasets, 1–3. The five trained weights have shown different sensitivity, specificity, accuracy, precision, and F1-Score measurements within the three datasets. Ensemble learning has shown that the Xception achieved a sensitivity of 90.91%, a specificity of 68.42%, an accuracy of 80.49%, a precision of 76.92%, and an F1-score of 83.33% for dataset 1. For dataset 2, Xception achieved a sensitivity of 87.50%, a specificity of 70.59%, an accuracy of 80.49%, a precision of 80.77%, and an F1-score of 84.00%. It has also shown that a sensitivity of 85.00%, a specificity of 57.14%, an accuracy of 70.73%, a precision of 65.38%, and an F1-score of 73.91% were achieved for dataset 3.

In Table 11, we demonstrated the ResNet50 performances of specified weights with ensemble learning on three RCFV datasets, 1–3. The five trained weights have shown different sensitivity, specificity, accuracy, precision, and F1-Score measurements within the three datasets. Ensemble learning has shown that the ResNet50 achieved a sensitivity of 73.08%, a specificity of 53.33%, an accuracy of 65.85%, a precision of 73.08%, and an F1-score of 73.08% for dataset 1. For dataset 2, the ResNet50 achieved a sensitivity of 100.00%, a specificity of 75.00%, an accuracy of 87.80%, a precision of 80.77%, and an F1-score of 89.36%. It has also shown that a sensitivity of 81.48%, a specificity of 71.43%, an accuracy of 78.05%, a precision of 84.62%, and an F1-score of 83.02% were achieved for dataset 3.



**Table 9.** The InceptionV3 performances of specified weights with ensemble learning.

RCFV Dataset	InceptionV3 Trained Weights	Sensitivity (%)	Specificity (%)	Accuracy (%)	Precision (%)	F1-Score (%)
1	10-epoch	88.00	75.00	82.93	84.62	86.27
	20-epoch	82.14	76.92	80.49	88.46	85.19
	30-epoch	82.14	76.92	80.49	88.46	85.19
	40-epoch	94.12	58.33	73.17	61.54	74.42
	50-epoch	76.47	100.00	80.49	100.00	86.67
	<b>Ensemble learning</b>	<b>85.71</b>	<b>84.62</b>	<b>85.37</b>	<b>92.31</b>	<b>88.89</b>
2	10-epoch	85.71	84.62	85.37	92.31	88.89
	20-epoch	85.71	84.62	85.37	92.31	88.89
	30-epoch	88.46	80.00	85.37	88.46	88.46
	40-epoch	76.47	100.00	80.49	100.00	86.67
	50-epoch	82.14	76.92	80.49	88.46	85.19
	<b>Ensemble learning</b>	<b>88.89</b>	<b>85.71</b>	<b>87.80</b>	<b>92.31</b>	<b>90.57</b>
3	10-epoch	74.19	70.00	73.17	88.46	80.70
	20-epoch	72.73	75.00	73.17	92.31	81.36
	30-epoch	68.42	100.00	70.73	100.00	81.25
	40-epoch	73.53	85.71	75.61	96.15	83.33
	50-epoch	73.53	85.71	75.61	96.15	83.33
	<b>Ensemble learning</b>	<b>74.29</b>	<b>100.00</b>	<b>78.05</b>	<b>100.00</b>	<b>85.25</b>

**Table 10.** The Xception performance of specified weights with ensemble learning.

RCFV Dataset	Xception Trained Weights	Sensitivity (%)	Specificity (%)	Accuracy (%)	Precision (%)	F1-Score (%)
1	10-epoch	94.12	58.33	73.17	61.54	74.42
	20-epoch	94.44	60.87	75.61	65.38	77.27
	30-epoch	84.21	54.55	68.29	61.54	71.11
	40-epoch	77.78	64.29	73.17	80.77	79.25
	50-epoch	71.88	66.67	70.73	88.46	79.31
	<b>Ensemble learning</b>	<b>90.91</b>	<b>68.42</b>	<b>80.49</b>	<b>76.92</b>	<b>83.33</b>
2	10-epoch	91.67	76.47	85.37	84.62	88.00
	20-epoch	90.48	65.00	78.05	73.08	80.85
	30-epoch	88.00	75.00	82.93	84.62	86.27
	40-epoch	85.19	78.57	82.93	88.46	86.79
	50-epoch	82.14	76.92	80.49	88.46	85.19
	<b>Ensemble learning</b>	<b>87.50</b>	<b>70.59</b>	<b>80.49</b>	<b>80.77</b>	<b>84.00</b>
3	10-epoch	81.82	57.89	70.73	69.23	75.00
	20-epoch	88.89	56.52	70.73	61.54	72.73
	30-epoch	85.71	60.00	73.17	69.23	76.60
	40-epoch	81.82	57.89	70.73	69.23	75.00
	50-epoch	76.00	56.25	68.29	73.08	74.51
	<b>Ensemble learning</b>	<b>85.00</b>	<b>57.14</b>	<b>70.73</b>	<b>65.38</b>	<b>73.91</b>

**Table 11.** The ResNet50 performances of specified weights with ensemble learning.

RCFV Dataset	ResNet50 Trained Weights	Sensitivity (%)	Specificity (%)	Accuracy (%)	Precision (%)	F1-Score (%)
1	10-epoch	65.79	66.67	65.85	96.15	78.13
	20-epoch	75.00	61.54	70.73	80.77	77.78
	30-epoch	73.08	53.33	65.85	73.08	73.08
	40-epoch	70.59	41.67	53.66	46.15	55.81
	50-epoch	71.43	53.85	65.85	76.92	74.07
	<b>Ensemble learning</b>	<b>73.08</b>	<b>53.33</b>	<b>65.85</b>	<b>73.08</b>	<b>73.08</b>
2	10-epoch	100.00	55.56	70.73	53.85	70.00
	20-epoch	86.96	66.67	78.05	76.92	81.63
	30-epoch	88.89	85.71	87.80	92.31	90.57
	40-epoch	100.00	60.00	75.61	61.54	76.19
	50-epoch	95.65	77.78	87.80	84.62	89.80
	<b>Ensemble learning</b>	<b>100.00</b>	<b>75.00</b>	<b>87.80</b>	<b>80.77</b>	<b>89.36</b>
3	10-epoch	75.00	61.54	70.73	80.77	77.78
	20-epoch	73.33	70.73	63.64	84.62	78.57
	30-epoch	86.36	63.16	75.61	73.08	79.17
	40-epoch	75.00	77.78	75.61	92.31	82.76
	50-epoch	73.53	85.71	75.61	96.15	83.33
	<b>Ensemble learning</b>	<b>81.48</b>	<b>71.43</b>	<b>78.05</b>	<b>84.62</b>	<b>83.02</b>

In Table 12, we demonstrated the multi-model weighted-averaging ensemble results using the five models' independent average-weighted ensemble performances from Table 7 to Table 11. Therefore, we compared single-model ensemble learning with multi-model ensemble learning. The performances in Table 12 show that multi-model ensemble learning achieved the same detection accuracy of 82.93% for all datasets. Therefore, this approach did not outperform the model when applied individually. Comparing ensemble learning in individual and combined results shows that the CheXNet model outperformed others and investigation-1.

**Table 12.** Final ensemble learning uses multi-weighted DL models on three different datasets.

RCFV Dataset	Ensemble of Models	Sensitivity (%)	Specificity (%)	Accuracy (%)	Precision (%)	F1-Score (%)
1	CheXNet, DenseNet,	88.00	75.00	82.93	84.62	86.27
2	InceptionV3, Xception,	88.00	75.00	82.93	84.62	86.27
3	Resnet50	80.65	90.00	82.93	96.15	87.72

#### 4.3. In Investigation-3

We calculated the true positive, true negative, false positive, and false negative values using the prediction label of each image from dataset A and dataset B, as demonstrated in Figure 6. Then, the performance of the five DL models was evaluated individually with sensitivity, specificity, accuracy, precision, and F1-score, indicating the percentage to which the model correctly identified both normal and pneumoconiosis CXRs. The individual and ensemble performances of the proposed models, CheXNet, DenseNet, InceptionV3, Xception, and ResNet50, are shown in Table 13. The LOOCV method was applied to find the most efficient model with the entire dataset.

**Table 13.** The performance of the leave-one-out method with five DL models.

Dataset	Efficiency Measurement	Model	Sensitivity (%)	Specificity (%)	Accuracy (%)	Precision (%)	F1-Score (%)
Contains 153 CXRs, including 71 Pneumoconiosis	Individually	CheXNet	<b>88.51</b>	<b>92.42</b>	<b>90.20</b>	<b>93.90</b>	<b>91.12</b>
		DenseNet	88.89	86.11	87.58	87.80	88.34
		InceptionV3	87.06	88.24	87.58	90.24	88.62
		Xception	85.88	86.76	86.27	89.02	87.43
		Resnet50	82.76	84.85	83.66	87.80	85.21
	Ensemble of five model's predictions	<b>90.14</b>	<b>92.68</b>	<b>91.50</b>	<b>91.43</b>	<b>90.78</b>	

Table 13 demonstrates that the proposed ensemble learning achieved the best performances on our dataset. As the most efficient method, CheXNet achieved the maximum accuracy of 90.20%, a sensitivity of 88.51%, specificity of 92.42%, a precision of 93.90%, and an F1-score of 91.12%. The ResNet50 model had the worst performance, and the other models' performances were not bad. Finally, the proposed ensemble achieved an accuracy of 91.50%, a sensitivity of 90.14%, a specificity of 92.68%, a precision of 91.43%, and an F1-score of 90.78% in our dataset.

### 5. Discussion

From investigation-1 to investigation-3, we applied different methodologies to improve pneumoconiosis detections in CXRs. In Table 14, we summarised the best statistical combination derived from the investigated ensemble techniques. Here, the lower the standard deviation (SD), the closer the values are to the mean of the set of investigations. The higher the SD, the wider the range of investigations. All techniques were processed to find the optimal solution for detecting pneumoconiosis from X-ray images. Investigation-1 had the best combination of accuracy of 82.93%, a sensitivity of 88.00%, a specificity of 75.00%, a precision of 84.62%, and an F1-score 86.27%, as summarised in Table 14, even though these are lower values than the individual performances of the CheXNet model without ensemble learning, as shown in Table 2. When compared to an individual, the performance of the ensemble learning technique in the first investigation did not improve pneumoconiosis detection's accuracy.

**Table 14.** Summary of best statistical combination achieved using the proposed techniques.

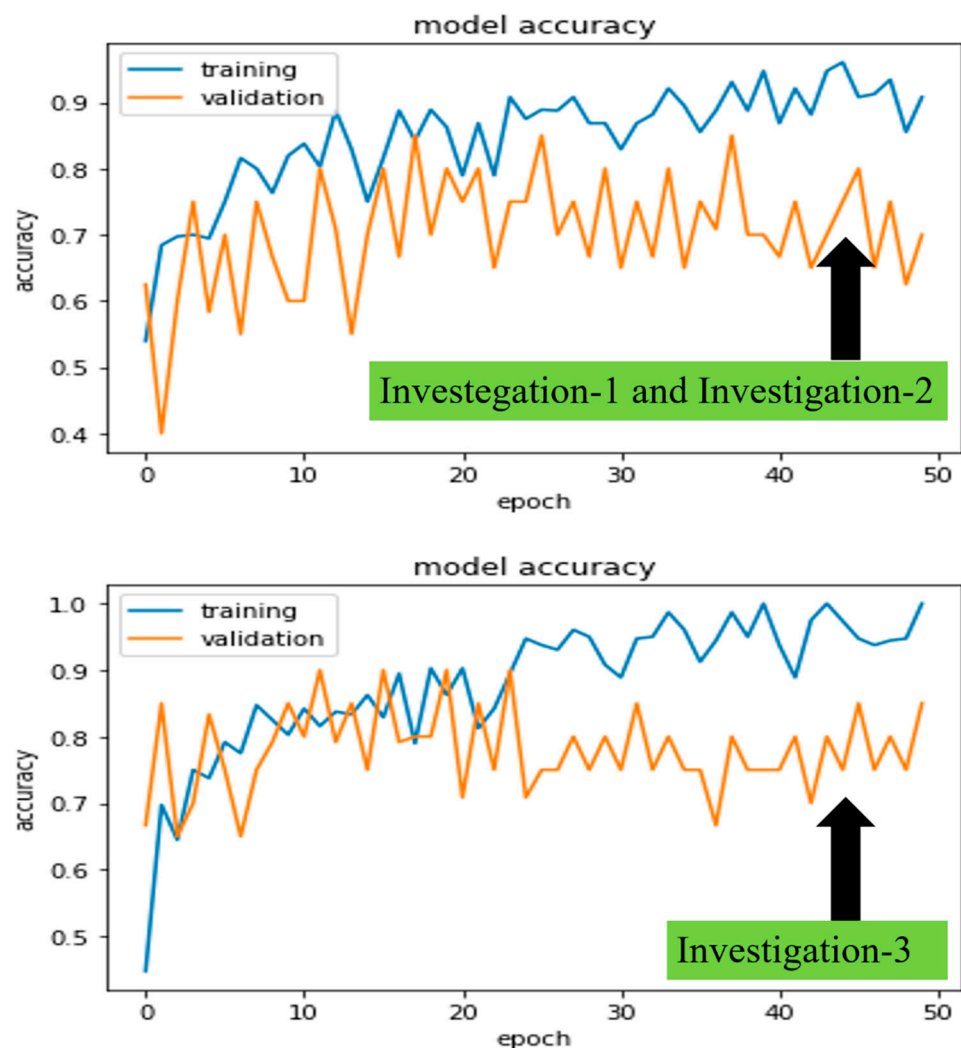
Techniques	Sensitivity (%)	Specificity (%)	Accuracy (%)	Precision (%)	F1-Score (%)
Investigation-1	88.00	75.00	82.93	84.62	86.27
Investigation-2	86.21	91.67	87.80	96.15	90.91
Investigation-3	<b>90.14</b>	<b>92.68</b>	<b>91.50</b>	<b>91.43</b>	<b>90.78</b>
Mean	88.12	86.45	87.41	90.73	89.32
SD	1.61	8.11	3.51	4.73	2.16

In investigation-2, we found that the detection performances slightly improved in the ensemble of multi-weighted averaging on a single model, CheXNet, as demonstrated in Table 14, which has shown improved statistical combinations than the other methodological findings in investigation-1. In investigation-3, we first observed that the same CheXNet model independently improved the accuracy from 87.80 to 90.20%. In addition, the proposed ensemble learning obtained 91.50% peak performance for detecting pneumoconiosis in coal workers from CXRs with state-of-the-art methods. Investigation-3 had an excellent success rate of more than 90.00% for all five observations. Therefore, as ground truth, our proposed ensemble learning outperformed other state-of-the-art classical and traditional machine and deep learning methods, as summarised in Table 1.

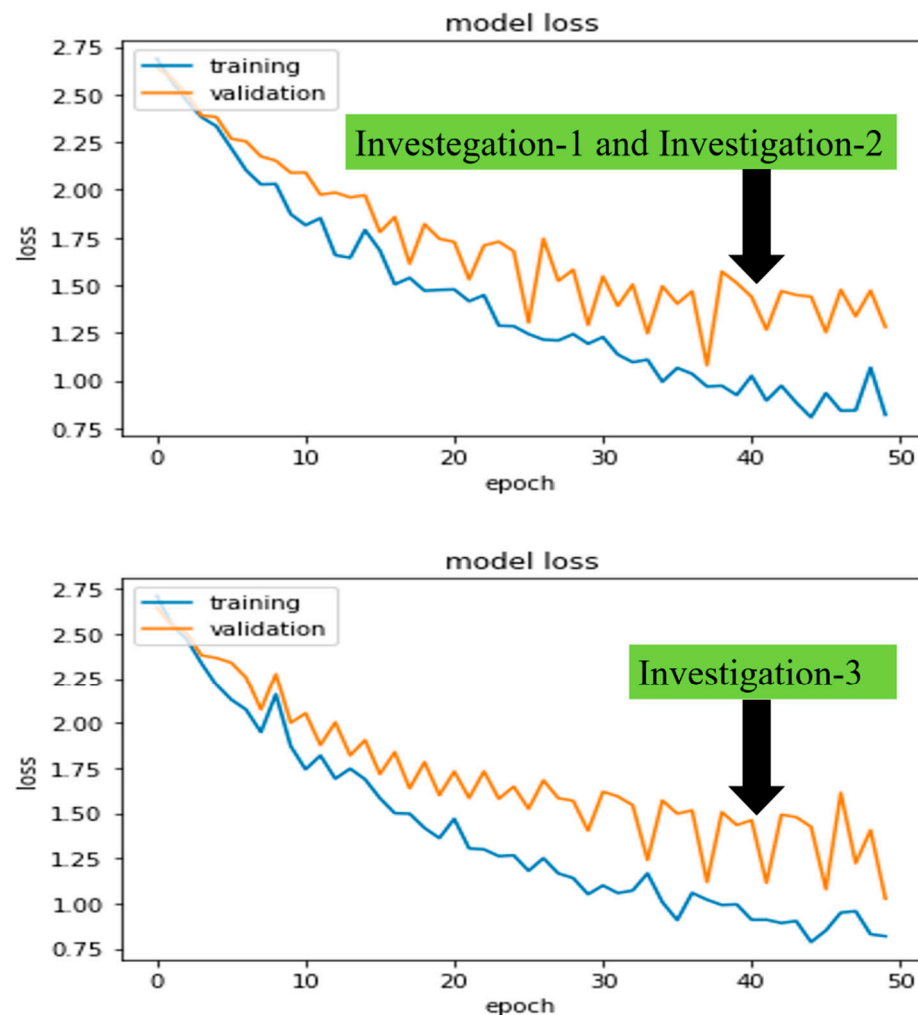
The University of Newcastle’s (Australia) high-performance computing (HPC) system was used for all investigations. Python 3.6 was used to run deep learning platform Keras 2.2.2 and machine learning platform Scikit-learn 0.19.1. In addition, we also looked at how long it took to train five different models, CheXNet, DenseNet, Inception-V3, Xception, and ResNet50, which took 19, 20, 16, 13, and 11 min, respectively, for 50 epochs. Furthermore, the model training and validation performance were monitored from the continuity of average (*Avg*) and standard deviation (*SD*) of accuracies and losses on each epoch. Figures 7 and 8 demonstrated the robust model, CheXNet training, and validation accuracies and losses for all proposed investigations. The investigated DL model was validated using Equations (3) and (4), where *i* indicates the *i*<sup>th</sup> epoch’s (*N* = 1 to 50) accuracy or loss values of a trained model. By comparing *Avg* and *SD*, we were able to pick the best-trained model to perform the test. In the following paragraphs, we present these values for the same robust model, CheXNet.

$$Avg = \frac{\sum_{i=1}^N i}{N} \tag{3}$$

$$SD = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (i - Avg)^2} \tag{4}$$



**Figure 7.** The CheXNet model’s training and validation accuracies in investigation-1 and 2 (on top) and investigation-3 (on bottom) per epoch.



**Figure 8.** The CheXNet model's training and validation losses in investigation-1 and 2 (**top**) and investigation-3 (**bottom**) per epoch.

In investigations 1 and 2, the *Avg* and *SD* of training and validation accuracies were calculated as  $Avg_{\text{Training}} = 0.86$ ,  $Avg_{\text{validation}} = 0.75$ ,  $SD_{\text{training}} = 0.09$ , and  $SD_{\text{validation}} = 0.05$  approximately. Similarly, losses were calculated as  $Avg_{\text{Training}} = 0.78$ ,  $Avg_{\text{validation}} = 0.97$ ,  $SD_{\text{training}} = 0.39$ , and  $SD_{\text{validation}} = 0.24$  approximately.

Finally, in investigation 3, the *Avg* and *SD* of training and validation accuracies were calculated as  $Avg_{\text{Training}} = 0.88$ ,  $Avg_{\text{validation}} = 0.78$ ,  $SD_{\text{training}} = 0.12$ , and  $SD_{\text{validation}} = 0.08$  approximately. Similarly, losses were calculated as  $Avg_{\text{Training}} = 0.7$ ,  $Avg_{\text{validation}} = 0.91$ ,  $SD_{\text{training}} = 0.33$ , and  $SD_{\text{validation}} = 0.18$  approximately.

The de-identified private CXRs database was gathered from the Coal Services Health NSW, St Vincent's Hospital, Sydney, Wesley Medical Imaging, Queensland, and ILO standard, which are supposed to comprise 100% correct assumptions for this research study. However, our proposed ensemble technique achieved an accuracy of 91.50%, a true positive rate (sensitivity) of 90.14%, and a true negative rate (specificity) of 92.68%, which were, on average, 10% lower than our assumptions.

This research study has a few limitations as well. First and foremost, the CSIRO's Sydney, Australia, office anonymised this private dataset, which cannot be accessed without their written consent [77]. However, if the proposed dataset is large, the outperformed ensemble investigation-3 may be mathematically expensive and take longer to obtain a robust assessment than other investigations. Future studies will focus on testing the proposed model in a clinical setting and gathering input to improve the methodology

further. Furthermore, we also recommend some form of variations in tool coupling to at least retain the best case.

## 6. Conclusions

In the paper, deep ensemble learning techniques were applied to detect pneumoconiosis automatically in the CXRs of coal workers. The ensemble was exploited by analysing the average probability, multi-weighted averaging, and majority label predictions using five deep learning models by using randomised cross-folds and leave-one-out cross-validations datasets. Three investigated results indicate the most efficient model, CheXNet on our small dataset that improves the accuracy from 85.37 to 90.20% independently. The integrated ensemble techniques with deep learning methods outperformed others, achieving an accuracy of 91.50% in the automated detection of pneumoconiosis. This study can be beneficial to researchers in the computer-aided diagnostic (CAD) system and to researchers dealing with small datasets in a real-time environment. Moreover, these investigations are useful for locating a reliable approach among the numerous alternatives. The approach substantially impacts clinical studies and is significant to physicians and other healthcare professionals.

**Author Contributions:** Conceptualisation, L.D. and K.S.; data curation, L.D. and K.S.; formal analysis, L.D., P.S., S.L., D.W., K.S., I.A.H. and F.S.A.; software, L.D. and K.S.; investigation, L.D., P.S., S.L., D.W., K.S., I.A.H. and F.S.A.; methodology, L.D., P.S., S.L., D.W., K.S., I.A.H. and F.S.A.; visualisation, L.D.; resources, L.D., P.S., S.L., D.W., K.S., I.A.H. and F.S.A.; writing—original draft, L.D., K.S., S.L., D.W. and P.S.; writing—review and editing, L.D., P.S., S.L., D.W., K.S., I.A.H. and F.S.A.; funding acquisition, K.S., I.A.H. and F.S.A.; supervision, S.L., P.S., D.W., I.A.H. and F.S.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** We would like to thank Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R319), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

**Acknowledgments:** We are grateful to CSIRO (Commonwealth Scientific and Industrial Research Organisation) Data61 for providing this study's segmented lung dataset. This higher-degree research collaboration has received partial financial support from the Coal Services Health and Safety Trust (Australia), Project #20647. We are grateful to all coal workers for allowing the use of their lungs as part of this study. We would like to thank Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R319), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Dietterich, T.G. Ensemble methods in machine learning. In *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*; LNCS; Springer: Milano, Italy, 2000; Volume 1857, pp. 1–15. [\[CrossRef\]](#)
2. Rajaraman, S.; Kim, I.; Antani, S.K. Detection and visualization of abnormality in chest radiographs using modality-specific convolutional neural network ensembles. *PeerJ* **2020**, *8*, e8693. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Rajaraman, S.; Sornapudi, S.; Kohli, M.; Antani, S. Assessment of an ensemble of machine learning models toward abnormality detection in chest radiographs. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBS, Berlin, Germany, 23–27 July 2019; pp. 3689–3692. [\[CrossRef\]](#)
4. Rajaraman, S.; Sornapudi, S.; Alderson, P.O.; Folio, L.R.; Antani, S.K. Analyzing inter-reader variability affecting deep ensemble learning for COVID-19 detection in chest radiographs. *PLoS ONE* **2020**, *15*, e0242301. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Rajaraman, S.; Jaeger, S.; Antani, S.K. Performance evaluation of deep neural ensembles toward malaria parasite detection in thin-blood smear images. *PeerJ* **2019**, *7*, e6977. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Izmailov, P.; Podoprikin, D.; Garipov, T.; Vetrov, D.; Wilson, A.G. Averaging Weights Leads to Wider Optima and Better Generalization. In Proceedings of the 34th Conference on Uncertainty in Artificial Intelligence 2018, UAI 2018, Monterey, CA, USA, 6–10 August 2018; Volume 2, pp. 876–885.
7. Tarvainen, A.; Valpola, H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 1195–1204.
8. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015—Conference Track Proceedings, San Diego, CA, USA, 7–9 May 2015; pp. 1–15.



9. Polyak, B.T.; Juditsky, A.B. Acceleration of Stochastic Approximation by Averaging. *SIAM J. Control Optim.* **1992**, *30*, 838–855. [[CrossRef](#)]
10. Ruppert, D. *Efficient Estimations from a Slowly Convergent Robbins-Monro Process*; Cornell University Operations Research and Industrial Engineering: Ithaca, NY, USA, 1988.
11. Lin, C.-W.; Wen, T.-C.; Setiawan, F. Evaluation of Vertical Ground Reaction Forces Pattern Visualization in Neurodegenerative Diseases Identification Using Deep Learning and Recurrence Plot Image Feature Extraction. *Sensors* **2020**, *20*, 3857. [[CrossRef](#)]
12. Khatamino, P.; Canturk, I.; Ozyilmaz, L. A Deep Learning-CNN Based System for Medical Diagnosis: An Application on Parkinson's Disease Handwriting Drawings. In Proceedings of the 2018 6th International Conference on Control Engineering & Information Technology (CEIT), Istanbul, Turkey, 25–27 October 2018. [[CrossRef](#)]
13. Rajaraman, S.; Siegelman, J.; Alderson, P.O.; Folio, L.S.; Folio, L.R.; Antani, S.K. Iteratively Pruned Deep Learning Ensembles for COVID-19 Detection in Chest X-Rays. *IEEE Access* **2020**, *8*, 115041–115050. [[CrossRef](#)]
14. Rajaraman, S.; Antani, S.K. Modality-Specific Deep Learning Model Ensembles Toward Improving TB Detection in Chest Radiographs. *IEEE Access* **2020**, *8*, 27318–27326. [[CrossRef](#)]
15. Rajaraman, S.; Sornapudi, S.; Alderson, P.O.; Folio, L.R.; Antani, S.K. Interpreting Deep Ensemble Learning through Radiologist Annotations for COVID-19 Detection in Chest Radiographs. *medRxiv* **2020**. [[CrossRef](#)]
16. Rajaraman, S.; Cemir, S.; Xue, Z.; Alderson, P.; Thoma, G.; Antani, S. A Novel Stacked Model Ensemble for Improved TB Detection in Chest Radiographs. *Med. Imaging* **2019**, 1–26. [[CrossRef](#)]
17. Rajaraman, S.; Folio, L.; Dimperio, J.; Alderson, P.; Antani, S. Improved Semantic Segmentation of Tuberculosis—Consistent Findings in Chest X-rays Using Augmented Training of Modality-Specific U-Net Models with Weak Localizations. *Diagnostics* **2021**, *11*, 616. [[CrossRef](#)] [[PubMed](#)]
18. Sivaramakrishnan, R.; Antani, S.; Candemir, S.; Xue, Z.; Thoma, G.; Alderson, P.; Abuya, J.; Kohli, M. Comparing deep learning models for population screening using chest radiography. *Med. Imaging 2018 Comput. Aided Diagn.* **2018**, 10575, 105751E. [[CrossRef](#)]
19. Kundu, R.; Das, R.; Geem, Z.W.; Han, G.-T.; Sarkar, R. Pneumonia detection in chest X-ray images using an ensemble of deep learning models. *PLoS ONE* **2021**, *16*, e0256630. [[CrossRef](#)] [[PubMed](#)]
20. Lopez-Martin, M.; Nevado, A.; Carro, B. Detection of early stages of Alzheimer's disease based on MEG activity with a randomized convolutional neural network. *Artif. Intell. Med.* **2020**, *107*, 101924. [[CrossRef](#)]
21. Lopez-Martin, M.; Carro, B.; Sanchez-Esguevillas, A. IoT type-of-traffic forecasting method based on gradient boosting neural networks. *Futur. Gener. Comput. Syst.* **2019**, *105*, 331–345. [[CrossRef](#)]
22. Rajpurkar, P.; Irvin, J.; Zhu, K.; Yang, B.; Mehta, H.; Duan, T.; Ng, A.Y. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. *arXiv* **2017**, arXiv:1711.05225.
23. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
24. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9. [[CrossRef](#)]
25. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1800–1807. [[CrossRef](#)]
26. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [[CrossRef](#)]
27. Devnath, L.; Luo, S.; Summons, P.; Wang, D. Automated detection of pneumoconiosis with multilevel deep features learned from chest X-Ray radiographs. *Comput. Biol. Med.* **2020**, *129*, 104125. [[CrossRef](#)]
28. Pulmonary Opacities on Chest X-ray • LITFL • CCC Differential Diagnosis. Available online: <https://litfl.com/pulmonary-opacities-on-chest-x-ray/> (accessed on 2 July 2020).
29. Guidelines for the Use of the ILO International Classification of Radiographs of Pneumoconioses, Revised Edition 2011. Available online: [http://www.ilo.org/global/topics/safety-and-health-at-work/resources-library/publications/WCMS\\_168260/lang-en/index.htm](http://www.ilo.org/global/topics/safety-and-health-at-work/resources-library/publications/WCMS_168260/lang-en/index.htm) (accessed on 2 July 2020).
30. Monash Centre for Occupational and Environmental Health. *Review of Respiratory Component of the Coal Mine Workers' Health Scheme for the Queensland Department of Natural Resources and Mines Final Report*; University of Illinois at Chicago: Chicago, IL, USA, 2016.
31. Kruger, R.P.; Thompson, W.B.; Turner, A.F. Computer Diagnosis of Pneumoconiosis. *IEEE Trans. Syst. Man Cybern.* **1974**, *4*, 40–49. [[CrossRef](#)]
32. Turner, A.F.; Kruger, R.P. Automated computer screening of chest radiographs for pneumoconiosis. *Invest. Radiol.* **1976**, *11*, 258–266. [[CrossRef](#)]
33. Chen, X.; Hasegawa, J.-I.; Toriwaki, J.-I. Quantitative diagnosis of pneumoconiosis based on recognition of small rounded opacities in chest X-ray images. In Proceedings of the International Conference on Pattern Recognition, Rome, Italy, 14–17 November 1988; pp. 462–464. [[CrossRef](#)]
34. Hall, E.L.; Crawford, W.O.; Roberts, F.E. Computer Classification of Pneumoconiosis from Radiographs of Coal Workers. *IEEE Trans. Biomed. Eng.* **1975**, *BME-22*, 518–527. [[CrossRef](#)]

35. Jagoe, J.R.; A Paton, K. Reading chest radiographs for pneumoconiosis by computer. *Occup. Environ. Med.* **1975**, *32*, 267–272. [[CrossRef](#)] [[PubMed](#)]
36. Alam, T.M.; Kamran, S.; Waseem, A.K.; Ibrahim, A.H.; Latifa, A.A.; Muhammad, A.R.; Memoona, A.; Suhuai, L. An Efficient Deep Learning-Based Skin Cancer Classifier for an Imbalanced Dataset. *Diagnostics* **2022**, *12*, 2115. [[CrossRef](#)]
37. Jagoe, J.R.; Paton, K.A. Measurement of Pneumoconiosis by Computer. *IEEE Trans. Comput.* **1976**, *C-25*, 95–97. [[CrossRef](#)]
38. Kobatake, H.; Oh'ishi, K.; Miyamichi, J. Automatic diagnosis of pneumoconiosis by texture analysis of chest X-ray images. In Proceedings of the ICASSP '87. IEEE International Conference on Acoustics, Speech, and Signal Processing, Dallas, TX, USA, 6–9 April 1987; Volume 12, pp. 610–613. [[CrossRef](#)]
39. Katsuragawa, S.; Doi, K.; MacMahon, H.; Nakamori, N.; Sasaki, Y.; Fennessy, J.J. Quantitative computer-aided analysis of lung texture in chest radiographs. *RadioGraphics* **1990**, *10*, 257–269. [[CrossRef](#)]
40. Murray, V.; Pattichis, M.S.; Davis, H.; Barriga, E.S.; Soliz, P. Multiscale AM-FM analysis of pneumoconiosis x-ray images. In Proceedings of the International Conference on Image Processing, ICIP, Cairo, Egypt, 7–10 November 2009; pp. 4201–4204. [[CrossRef](#)]
41. Savol, A.M.; Li, C.C.; Hoy, R.J. Computer-aided recognition of small rounded pneumoconiosis opacities in chest X-rays. *IEEE Trans. Pattern Anal. Mach. Intell.* **1980**, *PAMI-2*, 479–482. [[CrossRef](#)]
42. Kondo, H. Computer Aided Diagnosis for Pneumoconiosis Radiographs Using Neural Network. *Int. Arch. Photogramm. Remote Sens.* **2000**, *XXXIII*, 453–458.
43. Kondo, H.; Kouda, T. Computer-aided Diagnosis for Pneumoconiosis Using Neural Network. *Int. J. Biomed. Soft Comput. Hum. Sci. Off. J. Biomed. Fuzzy Syst. Assoc.* **2001**, *7*, 13–18. [[CrossRef](#)]
44. Kouda, T.; Kondo, H. Automatic Detection of Interstitial Lung Disease using Neural Network. *Int. J. Fuzzy Log. Intell. Syst.* **2002**, *2*, 15–19. [[CrossRef](#)]
45. Kondo, H.; Kouda, T. Detection of pneumoconiosis rounded opacities using neural network. In Proceedings of the Annual Conference of the North American Fuzzy Information Processing Society—NAFIPS, Vancouver, BC, Canada, 25–28 July 2001; Volume 3, pp. 1581–1585. [[CrossRef](#)]
46. Ledley, R.S.; Huang, H.; Rotolo, L.S. A texture analysis method in classification of coal workers 'pneumoconiosis'. *Comput. Biol. Med.* **1975**, *5*, 53–67. [[CrossRef](#)]
47. Ibrar, M.; Muhammad, A.H.; Kamran, S.; Talha, M.A.; Khaldoon, S.K.; Ibrahim, A.H.; Hanan, A.; Suhuai, L.. A Machine Learning-Based Model for Stability Prediction of Decentralized Power Grid Linked with Renewable Energy Resources. *Wirel. Commun. Mobile Comput.* **2022**, *2022*, 2697303. [[CrossRef](#)]
48. Yu, P.; Zhao, J.; Xu, H.; Sun, X.; Mao, L. Computer aided detection for pneumoconiosis based on Co-occurrence matrices analysis. In Proceedings of the 2009 2nd International Conference on Biomedical Engineering and Informatics, Tianjin, China, 17–19 October 2009; pp. 1–4. [[CrossRef](#)]
49. Shabbir, S.; Asif, M.S.; Alam, T.M.; Ramzan, Z. Early Prediction of Malignant Mesothelioma: An Approach Towards Non-invasive Method. *Curr. Bioinform.* **2021**, *16*, 1257–1277. [[CrossRef](#)]
50. Xu, H.; Tao, X.; Sundararajan, R.; Yan, W.; Annangi, P.; Sun, X.; Mao, L. Computer Aided Detection for Pneumoconiosis Screening on Digital Chest Radiographs. In Proceedings of the Third International Workshop on Pulmonary Image Analysis, Beijing, China, 20 September 2010; pp. 129–138.
51. Sundararajan, R.; Xu, H.; Annangi, P.; Tao, X.; Sun, X.; Mao, L. A multiresolution support vector machine based algorithm for pneumoconiosis detection from chest radiographs. In Proceedings of the 2010 7th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Rotterdam, The Netherlands, 14–17 April 2010; pp. 1317–1320. [[CrossRef](#)]
52. Tariq, A.; Awan, M.J.; Alshudukhi, J.; Alam, T.M.; Alhamazani, K.T.; Meraf, Z. Software Measurement by Using Artificial Intelligence. *J. Nanomater.* **2022**, *2022*, 7283171. [[CrossRef](#)]
53. Alam, T.M.; Shaukat, K.; Khelifi, A.; Khan, W.A.; Raza, H.M.E.; Idrees, M.; Luo, S.; Hameed, I.A. Disease Diagnosis System Using IoT Empowered with Fuzzy Inference System. *Comput. Mater. Contin.* **2022**, *70*, 5305–5319.
54. Baig, T.I.; Khan, Y.D.; Alam, T.M.; Biswal, B.; Aljuaid, H.; Gillani, D.Q. ILipo-PseAAC: Identification of lipoylation sites using statistical moments and general PseAAC. *Comput. Mater. Contin.* **2022**, *71*, 215–230.
55. Shaukat, K.; Alam, T.M.; Hameed, I.A.; Luo, S.; Li, J.; Aujla, G.K.; Iqbal, F. A comprehensive dataset for bibliometric analysis of SARS and coronavirus impact on social sciences. *Data Brief* **2020**, *33*, 106520. [[CrossRef](#)]
56. Baig, T.I.; Alam, T.M.; Anjum, T.; Naseer, S.; Wahab, A.; Imtiaz, M.; Raza, M.M. Classification of human face: Asian and non-Asian people. In Proceedings of the 2019 International Conference on Innovative Computing (ICIC), Lahore, Pakistan, 1–2 November 2019; pp. 1–6.
57. Alam, T.M.; Shaukat, K.; Mahboob, H.; Sarwar, M.U.; Iqbal, F.; Nasir, A.; Hameed, I.A.; Luo, S. A Machine Learning Approach for Identification of Malignant Mesothelioma Etiological Factors in an Imbalanced Dataset. *Comput. J.* **2021**, *65*, 1740–1751. [[CrossRef](#)]
58. Latif, M.Z.; Shaukat, K.; Luo, S.; Hameed, I.A.; Iqbal, F.; Alam, T.M. Risk Factors Identification of Malignant Mesothelioma: A Data Mining Based Approach. In Proceedings of the 2020 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), Istanbul, Turkey, 12–13 June 2020; pp. 1–6.
59. Khushi, M.; Shaukat, K.; Alam, T.M.; Hameed, I.A.; Uddin, S.; Luo, S.; Yang, X.; Reyes, M.C. A Comparative Performance Analysis of Data Resampling Methods on Imbalance Medical Data. *IEEE Access* **2021**, *9*, 109960–109975. [[CrossRef](#)]

60. Alam, T.M.; Shaukat, K.; Hameed, I.A.; Khan, W.A.; Sarwar, M.U.; Iqbal, F.; Luo, S. A novel framework for prognostic factors identification of malignant mesothelioma through association rule mining. *Biomed. Signal Process. Control* **2021**, *68*, 102726. [[CrossRef](#)]
61. Shaukat, K.; Iqbal, F.; Alam, T.M.; Aujla, G.K.; Devnath, L.; Khan, A.G.; Rubab, A. The impact of artificial intelligence and robotics on the future employment opportunities. *Trends Comput. Sci. Inf. Technol.* **2020**, *5*, 50–54.
62. Ghani, M.U.; Alam, T.M.; Jaskani, F.H. Comparison of classification models for early prediction of breast cancer. In Proceedings of the 2019 International Conference on Innovative Computing (ICIC), Seoul, Korea, 26–29 August 2019; pp. 1–6.
63. Nasir, A.; Shaukat, K.; Hameed, I.A.; Luo, S.; Alam, T.M.; Iqbal, F. A Bibliometric Analysis of Corona Pandemic in Social Sciences: A Review of Influential Aspects and Conceptual Structure. *IEEE Access* **2020**, *8*, 133377–133402. [[CrossRef](#)]
64. Alam, T.M.; Khan, M.M.A.; Iqbal, M.A.; Abdul, W.; Mushtaq, M. Cervical cancer prediction through different screening methods using data mining. *Int. J. Adv. Comput. Sci. Appl.* **2019**, *10*. [[CrossRef](#)]
65. Ali, Y.; Farooq, A.; Alam, T.M.; Farooq, M.S.; Awan, M.J.; Baig, T.I. Detection of Schistosomiasis Factors Using Association Rule Mining. *IEEE Access* **2019**, *7*, 186108–186114. [[CrossRef](#)]
66. Alam, T.M.; Iqbal, M.A.; Ali, Y.; Wahab, A.; Ijaz, S.; Baig, T.I.; Hussain, A.; Malik, M.A.; Raza, M.M.; Ibrar, S.; et al. A model for early prediction of diabetes. *Inform. Med. Unlocked* **2019**, *16*, 100204. [[CrossRef](#)]
67. Rajaraman, S.; Antani, S. Visualizing Salient Network Activations in Convolutional Neural Networks for Medical Image Modality Classification. *Commun. Comput. Inf. Sci.* **2018**, *1036*, 42–57. [[CrossRef](#)]
68. Zhang, L.; Rong, R.; Li, Q.; Yang, D.M.; Yao, B.; Luo, D.; Zhang, X.; Zhu, X.; Luo, J.; Liu, Y.; et al. A deep learning-based model for screening and staging pneumoconiosis. *Sci. Rep.* **2021**, *11*, 2201. [[CrossRef](#)]
69. Rajaraman, S.; Silamut, K.; Hossain, A.; Ersoy, I.; Maude, R.J.; Jaeger, S.; Thoma, G.R.; Antani, S.K. Understanding the learned behavior of customized convolutional neural networks toward malaria parasite detection in thin blood smear images. *J. Med Imaging* **2018**, *5*, 034501. [[CrossRef](#)]
70. Thamizhvani, T.R.; Lakshmanan, S.; Sivaramakrishnan, R. Computer Aided Diagnosis of Skin Tumours from Dermal Images. *Lect. Notes Comput. Vis. Biomech.* **2018**, *28*, 349–365. [[CrossRef](#)]
71. Thamizhvani, T.R.; Lakshmanan, S.; Sivaramakrishnan, R. Mobile application-based computer-aided diagnosis of skin tumours from dermal images. *Imaging Sci. J.* **2018**, *66*, 382–391. [[CrossRef](#)]
72. Sivaramakrishnan, R.; Antani, S.; Jaeger, S. Visualising Deep Learning Activations for Improved Malaria Cell Classification. In Proceedings of the First Workshop Medical Informatics and Healthcare held with the 23rd SIGKDD Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 10–18 October 2017; pp. 40–47.
73. Devnath, L.; Luo, S.; Summons, P.; Wang, D. Tuberculosis (TB) Classification in Chest Radiographs using Deep Convolutional Neural Networks. *Int. J. Adv. Sci. Eng. Technol. IJASEAT* **2018**, *6* (Suppl. 1), 68–74.
74. Devnath, L.; Peter, S.; Suhuai, L.; Dadong, W.; Kamran, S.; Ibrahim, A.H.; Hanan, A. Computer-Aided Diagnosis of Coal Workers' Pneumoconiosis in Chest X-ray Radiographs Using Machine Learning: A Systematic Literature Review. *Int. J. Environ. Res. Public Health* **2022**, *19*, 6439. [[CrossRef](#)]
75. Wang, D.; Arzhaeva, Y.; Devnath, L.; Qiao, M.; Amirgholipour, S.; Liao, Q.; McBean, R.; Hillhouse, J.; Luo, S.; Meredith, D.; et al. Automated Pneumoconiosis Detection on Chest X-Rays Using Cascaded Learning with Real and Synthetic Radiographs. In Proceedings of the 2020 Digital Image Computing: Techniques and Applications (DICTA), Melbourne, Australia, 29 November–2 December 2020; pp. 1–6. [[CrossRef](#)]
76. Devnath, L.; Luo, S.; Summons, P.; Wang, D. An accurate black lung detection using transfer learning based on deep neural networks. In Proceedings of the International Conference Image and Vision Computing, Dunedin, New Zealand, 2–4 December 2019. [[CrossRef](#)]
77. Devnath, L.; Luo, S.; Summons, P.; Wang, D. Performance comparison of deep learning models for black lung detection on chest X-ray radiographs. In *ACM International Conference Proceeding Series*; ACM: Sydney, Australia, 2020; pp. 152–154. [[CrossRef](#)]
78. Arzhaeva, Y.; Wang, D.; Devnath, L.; Amirgholipour, S.K.; McBean, R.; Hillhouse, J.; Yates, D. Development of Automated Diagnostic Tools for Pneumoconiosis Detection from Chest X-ray Radiographs. In *The Final Report Prepared for Coal Services Health and Safety Trust*; Coal Services: Sydney, Australia, 2019.
79. Wang, X.; Yu, J.; Zhu, Q.; Li, S.; Zhao, Z.; Yang, B.; Pu, J. Potential of deep learning in assessing pneumoconiosis depicted on digital chest radiography. *Occup. Environ. Med.* **2020**, *77*, 597–602. [[CrossRef](#)] [[PubMed](#)]
80. Yang, F.; Tang, Z.-R.; Chen, J.; Tang, M.; Wang, S.; Qi, W.; Yao, C.; Yu, Y.; Guo, Y.; Yu, Z. Pneumoconiosis computer aided diagnosis system based on X-rays and deep learning. *BMC Med. Imaging* **2021**, *21*, 189. [[CrossRef](#)] [[PubMed](#)]
81. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. Published as a Conference Paper at ICLR. 2015. Available online: <https://arxiv.org/abs/1409.1556> (accessed on 26 April 2020).