Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (IJCAI-25)
Special Track on AI and Social Good

SHIELD: A Self-supervised, Silicosis-focused Hierarchical Imaging Framework for Occupational Lung Disease Diagnosis

Yasmeena Akhter , Rishabh Ranjan , Richa Singh , Mayank Vatsa
Indian Institute of Technology Jodhpur, India
{akhter.1, ranjan.4, richa, mvatsa}@iitj.ac.in

Abstract

Silicosis is an irreversible lung disease caused by silica dust exposure in industrial settings. Early detection is crucial, but automatic diagnostic methods are hindered by limited data availability. We propose *SHIELD* - a self-supervised, *S*ilicosis-focused *H*ierarchical *I*maging framework for early occupational *L*ung disease *D*iagnosis. Our method leverages a multi-resolution jigsaw puzzle pretext task on CXR images to extract and preserve features for lung region analysis. By employing a pyramidal strategy to generate pretrained models at various resolutions, followed by fine-tuning and a two-level ensembling across diverse deep learning architectures, *SHIELD* achieves enhanced diagnostic accuracy. We validate our approach on a publicly collected CXR dataset of *3044* samples from public health centers in India. *SHIELD* achieves 72% accuracy, demonstrating up to 20% improvement over baseline approaches. This work advances medical image analysis and supports UN Sustainable Development Goal 3 by providing cost-effective early screening in resource-limited settings.

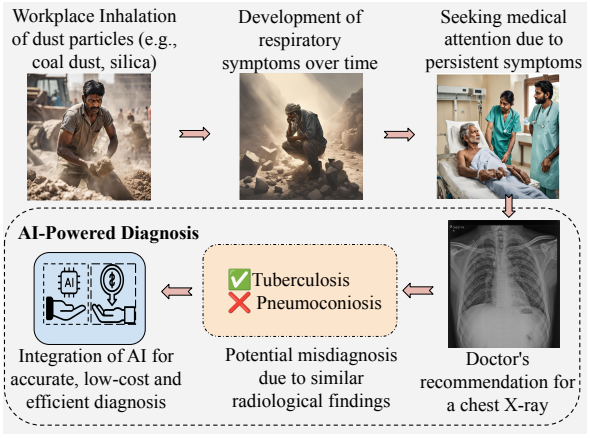

Figure 1: Visual representation of Silicosis progression and diagnostic challenges in resource-constrained healthcare settings. The disease's radiological similarities with Tuberculosis often lead to misdiagnosis, particularly in high-volume centers. SHIELD addresses this through AI-driven differential diagnosis, enabling accurate early detection and intervention.

1 Introduction

Pneumoconiosis is a chronic occupational lung disease characterized by pulmonary interstitial fibrosis, resulting from prolonged exposure to industrial dust particles. The condition develops through sustained inhalation of inorganic materials, including silica, coal, asbestos, beryllium, and cobalt, primarily affecting workers in mining, construction, and manufacturing sectors. The microscopic particles progressively accumulate in lung tissue, causing irreversible scarring and respiratory dysfunction. *Silicosis*, a severe form of Pneumoconiosis, is caused by prolonged inhalation of crystalline silica (SiO_2) dust. This progressive and irreversible condition predominantly affects workers in high-risk industries, including stone carving, mining, construction, ceramics, glass manufacturing, and sandblasting operations. The disease's prevalence and severity make it a significant occupational health concern, particularly in developing regions where workplace safety measures may be limited. Global cases have risen by 61.5% from 1990 to 2019 [Huang *et al.*, 2023b], with approximately 99% of pollution-related deaths occurring in developing countries [World Health Organization, 2021; Barnes *et al.*, 2019; Mandrioli *et al.*, 2018]. In 2019 alone, China reported 136,755 cases, India documented 11,670 cases, and the United States recorded 10,014 cases [Momtazmanesh *et al.*, 2023], representing a public health challenge [Akhter *et al.*, 2023a].

The technical complexity of Silicosis diagnosis stems from multiple challenges in medical image analysis. The disease's irreversible nature demands early detection [Davies, 1957; Leung *et al.*, 2012], requiring accurate identification of subtle radiological patterns. However, these patterns often mimic other respiratory conditions such as Tuberculosis [Cowie, 1994; Ehrlich *et al.*, 2006; Ehrlich *et al.*, 2021], creating a complex pattern recognition problem. Traditional diagnostic approaches face significant limitations in scaling to meet the needs of millions of at-risk workers, particularly given the shortage of specialist radiologists and the economic burden of healthcare expenses, which push nearly 100 million

people into poverty annually [Neelsen *et al.*, 2022]. These challenges necessitate automated, scalable solutions that can maintain diagnostic accuracy while processing large volumes of CXRs. Figure 1 illustrates this critical pathway from occupational exposure to diagnosis, highlighting the key challenge of distinguishing Silicosis from similar pulmonary conditions such as Tuberculosis and demonstrating how AI-powered solutions can support accurate, efficient diagnosis in resource-constrained settings.

This technical challenge directly impacts multiple UN Sustainable Development Goals (SDGs) and the "Leave No One Behind" commitment, requiring innovative solutions that balance computational burden with practical deployment constraints. The development of accurate, automated diagnostic systems could advance progress toward ensuring healthy lives (SDG 3), decent work conditions (SDG 8), and reducing inequalities (SDG 10), particularly in resource-constrained settings where specialist medical expertise is limited. The need for scalable, accurate, and accessible screening solutions points toward Artificial Intelligence-based approaches, specifically deep learning techniques that can learn complex radiological patterns while maintaining robustness across diverse healthcare settings.

1.1 Related Work

While deep learning has revolutionized medical imaging diagnosis [Mirbabaie *et al.*, 2021; Alowais *et al.*, 2023; Göndöcs and Dörfler, 2024; Eisemann *et al.*, 2025], its application to Pneumoconiosis detection in general faces unique challenges. Earlier Computer-Aided Diagnosis (CAD) attempts [Okumura *et al.*, 2011; Zhu *et al.*, 2014a; Zhu *et al.*, 2014b; Okumura *et al.*, 2017; Pattichis *et al.*, 2002; Sundararajan *et al.*, 2010; Yu *et al.*, 2011] relied on limited in-house datasets, a challenge that persists in recent research as well. Current studies remain constrained by small sample sizes, with datasets ranging from 700 to 1,760 samples [Yang *et al.*, 2021; Zhang *et al.*, 2021], while similar limitations appear in other works [Xu *et al.*, 2010; Wang *et al.*, 2020; Devnath *et al.*, 2021; Wang *et al.*, 2021]. Researchers [Yang *et al.*, 2021] developed a two-stage pipeline using UNet for lung segmentation and ResNet34 for classification, validated on 1,760 CXR images. [Zhang *et al.*, 2021] introduced a region-based approach, dividing CXRs into six subregions for opacity detection and 4-class staging, tested on 1,216 subjects. [Devnath *et al.*, 2021] leveraged transfer learning with CheXNet [Rajpurkar *et al.*, 2017], combining multilevel features with SVM configurations. [Wang *et al.*, 2021] evaluated InceptionV3 [Szegedy *et al.*, 2015] on 1,881 samples, while [Wang *et al.*, 2020] proposed a cascaded framework using CycleGAN for synthetic data generation.

These constraints stem from the complexity of radiological annotation and privacy concerns [Akhter *et al.*, 2023b; Mushtaq *et al.*, 2024]. Self-Supervised Learning (SSL) has emerged as a promising solution [Huang *et al.*, 2023a; Haghighi *et al.*, 2024; Manna *et al.*, 2024; Azizi *et al.*, 2021; Zhou *et al.*, 2019], offering the ability to leverage unlabeled data for pretraining before fine-tuning on smaller datasets. To overcome the challenge, we leverage SSL, a paradigm that enhances model robustness while reducing dependence on labeled data. While [Zhang *et al.*, 2021] demonstrated success using sub-region-based classification aligned with ILO standards, we propose a novel SSL framework utilizing multi-resolution jigsaw puzzle solving as a pretext task. Our approach preserves global anatomical context and spatial relationships critical for pneumoconiosis diagnosis, while eliminating the need for granular sub-region annotations. This end-to-end learning strategy enables the model to learn hierarchical features directly from the whole CXR, improving both efficiency and diagnostic accuracy.

1.2 Research Contributions

Pneumoconiosis, a group of occupational lung diseases, particularly Silicosis, presents significant healthcare challenges in industrial settings worldwide, necessitating innovative diagnostic approaches. The complete workflow for the proposed work is shown in Fig.2. This multidisciplinary, collaborative approach to healthcare innovation and sustainable development makes the following key contributions:

1. **Scientific Innovation:** We propose *SHIELD*, a **S**elf-supervised, **S**ilicosis-focused **H**ierarchical **I**maging framework for early occupational **L**ung disease **D**iagnosis, utilizing multi-resolution jigsaw puzzle solving as a pretext task, combined with a two-level ensemble strategy for pneumoconiosis detection. This approach effectively addresses the *small sample space* challenge while preserving crucial spatial relationships in lung region analysis at different resolutions, advancing the state-of-the-art in medical image analysis.
2. **Experimental Validation and Societal Impact:** Validated on $3,000+$ CXRs from Silicosis cases across multiple healthcare centers, the proposed SHIELD method achieves 72% average accuracy across all classes, showing a 2% improvement over the best baseline (SSL-1024) and up to 20% improvement over traditional transfer learning approaches. Our approach enables early prediction, improving worker health outcomes and reducing economic burden.
3. **Multi-stakeholder Collaboration**: Our research brings together AI, medicine, and public health to develop an AI-enabled diagnostic system through close collaboration with healthcare professionals. Incorporating iterative stakeholder feedback, we designed a GUI-based tool tailored for public healthcare settings. Currently undergoing user interface testing in clinical environments, our approach ensures practical deployment in resource-limited settings while addressing real-world implementation challenges.

2 Methodology

Given a dataset $\mathcal{D} = (\mathbf{x}i, y_i)i = 1^N$ of CXR images $\mathbf{x}_i \in \mathbb{R}^{H \times W}$ and corresponding labels $y_i \in 0, \ldots, K$ where K represents distinct pathological conditions based classes including Silicosis, Silico Tuberculosis (STB), Tuberculosis (TB) and Normal, constituting a multiclass classification problem. Our objective is to learn a deep neural network

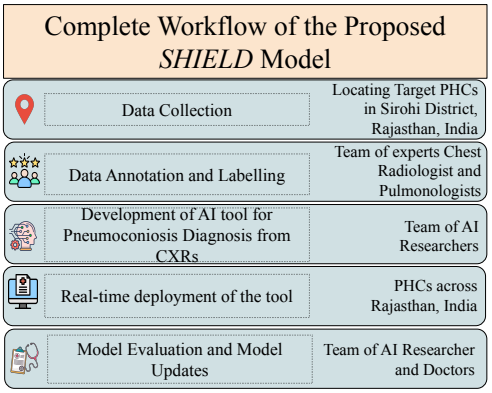

Figure 2: End-to-end workflow for *SHIELD*, from data collection to deployment. The system integrates clinical expertise with deep learning for automated Silicosis analysis.

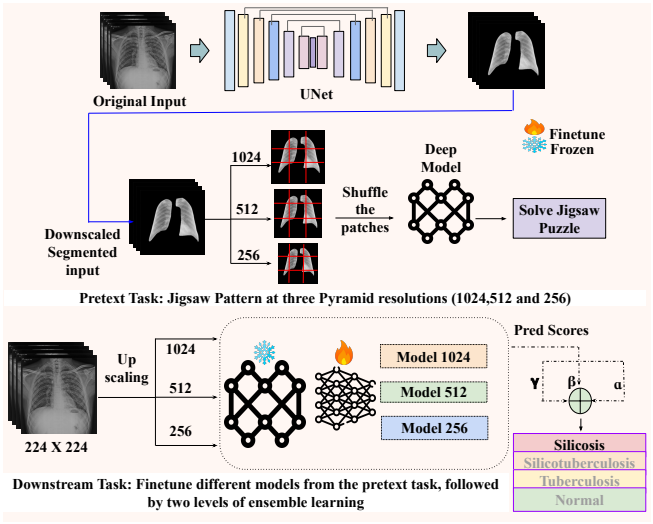

Figure 3: *SHIELD*: Illustrating deep-learning framework for diagnosing Silicosis through a two-phase approach. The first phase involves lung segmentation and self-supervised learning using a multi-resolution jigsaw puzzle strategy as the pretext task, while the second phase implements a downstream task comprising fine-tuning and ensemble learning to achieve accurate disease classification.

$f(\mathbf{x};\theta) : \mathbb{R}^{H\times W} \rightarrow \mathbb{R}^{K}$ that maps input images to class probabilities by minimizing the cross-entropy loss.

The limited availability of labeled medical data and the complex nature of radiological patterns necessitate an approach that can effectively learn robust features from limited samples while maintaining diagnostic accuracy across diverse clinical settings. Our proposed diagnostic framework, *SHIELD*, for early occupational Lung disease Diagnosis, consists of two main components: (1) a pretext task utilizing a multi-resolution jigsaw puzzle mechanism for SSL, and (2) a downstream task leveraging the learned representations for lung disease classification through a two-level ensemble learning mechanism. The framework is designed to effectively learn hierarchical features from CXRs, using different spatial resolutions with enhanced representational learning from *SSL* pretraining approach. The final classification is a two-stage ensemble boosted prediction outcome, combining the multi-resolution and weighted score fusion.

2.1 Pretext Task: Multi-Resolution Jigsaw Learning

The pretext task is structured as an SSL problem where we transform chest CXR images into jigsaw puzzles with multiple-resolution settings. This approach lets the model learn local and global context and information across different resolutions.

Preliminary Step: Image Preprocessing and Segmentation: Given an input CXR image $I \in \mathbb{R}^{H\times W}$, we apply lung segmentation to isolate lung fields, removing background noise and normalizing the input space. This produces a binary mask, *M*, which is multiplied with the original image to generate segmented inputs for the pretext task. This segmented representation serves as the foundation for subsequent multi-resolution analysis. To create the segmentation mask, we employed the UNet [Ronneberger *et al.*, 2015] architecture as the segmentation model, which is trained and validated on the Shenzhen and Montgomery datasets [Jaeger *et al.*, 2014] with ResNet50 [He *et al.*, 2016] backbone, and pretrained on the ImageNet [Krizhevsky *et al.*, 2012] dataset. The trained model is applied to Silicosis CXR samples, performing segmentation at three resolutions (1024, 512, 256), enabling focused multi-resolution analysis.

Multi-Resolution Pyramid Generation: We generate a pyramid of resolutions (1024 × 1024, 512 × 512, and 256 × 256) from the segmented lung image. Each resolution level L is represented as $I_L = R(I, S_L)$, where $L \in \{1024, 512, 256\}$ and R(·) is the resize operation and S_L is the target size at level *L*. This multi-resolution approach ensures that the model learns features at different spatial scales, crucial for capturing fine-grained details and global structural patterns.

Jigsaw Pattern Generation as Pretext Task: The self-supervised jigsaw puzzle pretext task is particularly well-suited for CXR analysis due to the inherent anatomical symmetry and structural relationships in thoracic imagery. By decomposing CXRs into patches and learning their spatial arrangements, the model learns critical anatomical relationships, tissue patterns, and organ boundaries fundamental to diagnostic interpretation. This pretext task is especially valuable as it enforces the model to understand both local features (e.g., tissue textures, anatomical patterns) and global structural relationships (e.g., cardiothoracic ratio, lung field symmetry) without requiring expert annotations. Furthermore, the multi-resolution approach in our jigsaw framework enables capturing features at varying scales, which is crucial for detecting diffuse diseases (like interstitial patterns) and localized abnormalities (such as nodules) in CXRs. The pretext tasks encourage learning more representational features based on the anatomical structure of the lungs. To improve this, we performed the pretext task at three resolutions. The resolution-specific pretraining ensures each model develops unique and complementary feature extraction capabilities, significantly enhancing the ensemble's discriminative

power. At each resolution level, we divide the image into a grid of 3×3 patches. The patches are shuffled according to a predefined permutation strategy P, creating a jigsaw puzzle using $J_l = P(I_L, 3 \times 3)$, where J_L represents the shuffled image at resolution *L*, the permutation P is selected from a set of valid permutations that maintain sufficient complexity while ensuring learning feasibility.

We adapted multiple deep learning-based computer vision models to solve the Jigsaw puzzle on the 3 × 3 puzzle pattern and finetuned each model individually. The choice for the models included ResNet18[He *et al.*, 2016], VGG16[Liu and Deng, 2015], AlexNet[Krizhevsky *et al.*, 2012] and DenseNet121[Huang *et al.*, 2017], pretrained on the ImageNet dataset. As mentioned above, the pretext task inputs the segmented CXR. The pretext task results in a total of 12 models across three resolutions for 4 different deep models.

2.2 Downstream Task: Multi-Model Ensemble Classification

In the downstream task, we leverage the rich hierarchical representations learned during the jigsaw puzzle pretraining phase (pretext task) to finetune the model for the specific task of lung disease classification, where the encoder's learned understanding of anatomical structures and spatial relationships are crucial for differentiating among various types of Pneumoconiosis such as Silicosis, TB, and their combinations (STB).

In our case, we performed the downstream task for classifying a given CXR image using the pretrained models from the pretext task at three resolutions by finetuning them at the resolution of 224 × 224, which is the standard resolution to train the deep learning based computer vision models, optimizing the computational cost as wel. In the finetuning stage, we rescaled the 224 × 224 to the multiscale pyramid level, L, where $L \in \{256, 512, 1024\}$, for all the four models. Each model is initialized with weights learned from the pretext task and finetuned for the classification task for Silicosis, STB, TB, and normal. Meanwhile, we selected the best-performing models across the three resolutions to generate the ensemble-boosted results. Thus, it paves the way for the first level of ensemble learning to improve classification results. The models differ in their architectural details to ensure diversity in feature extraction represented as $model_{1024}$ - optimized for high-resolution features, $model_{512}$ - balanced for mid-resolution features and $model_{256}$ - specialized for low-resolution features at a resolution of 1024 × 1024, 512 × 512 and 256 × 256, respectively. This is followed by the adaptive weighted score fusion of the prediction scores generated by the models across the three resolutions for optimizing ensemble results. The weighted score fusion is followed using the equation 1.

$$FinalPred = \alpha * Model_{1024} + \beta * Model_{512} + \gamma * Model_{256} \quad (1)$$

where α, β, and γ are learnable parameters to tune the equation 1 for the contribution of each model to the final prediction outcome. The two-level ensemble strategy encompasses the strengths of multi-solution and adaptive weighted score fusion. The complete pipeline of the proposed classification approach is shown in Fig. 3. This comprehensive methodology enables robust feature learning and accurate classification of CXR images into one of the aforementioned classes, leveraging both self-supervised and ensemble learning approaches.

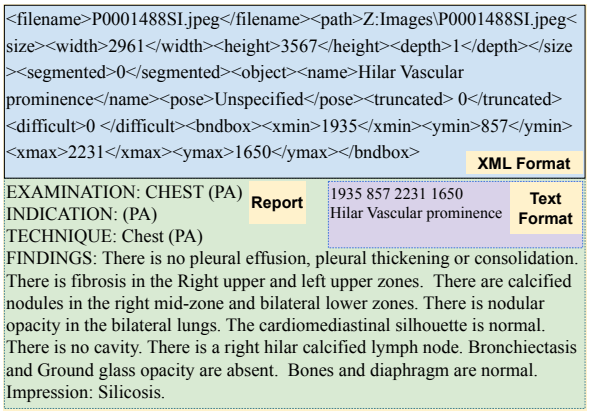

Figure 4: Showcases the annotation output of the collected dataset in XML, text, and the corresponding report.

3 Experimental Setup

Dataset: To develop *SHIELD*, we retrospectively collected a comprehensive CXR dataset [Akhter *et al.*, 2025], specifically curated for deep learning-based Pneumoconiosis diagnosis. The dataset comprises 3,044 high-resolution frontal chest radiographs collected from stone workers across multiple Primary health care centers in the Sirohi district in Rajasthan, India. Each grayscale image, with an average resolution of 3567 × 2898 pixels, was acquired using standard radiological protocols and stored in JPEG format. The data collection spanned three years, ensuring temporal diversity and real-world clinical variability. The dataset encompasses four distinct diagnostic categories: Silicosis (a progressive fibrotic lung disease), Silicotuberculosis (STB, a complex combination of silicosis and Tuberculosis), Tuberculosis (TB), and normal cases. Each image was annotated through a rigorous multi-stage process involving experienced radiologists, with inter-annotator agreement maintained above 0.85 (Cohen's kappa). To ensure annotation quality, we employed a consensus-based approach where disagreements were resolved through panel discussions.

Representative samples from the collected dataset are illustrated in Fig. 5 for each of the classes. The class distribution and train-test split protocol are detailed in Table 1, maintaining stratification across diagnostic categories to prevent class imbalance issues during model training. This carefully curated dataset forms the foundation of *SHIELD*, AI-based diagnostic framework aimed at enabling early, efficient, and cost-effective screening of occupational lung diseases.

Dataset Compilation: To annotate and label the collected dataset, we teamed up with a group of expert radiologists who performed the data labelling and annotation. The dataset annotation was conducted through a structured process involving five board-certified radiologists. To maintain annotation

Set	Silicosis	STB	TB	Normal	Total
Train	738	529	525	337	2129
Test	319	227	225	144	915
Total	1057	752	750	481	3044

Table 1: Classwise distribution of the compiled CXR dataset for Silicosis, used to assess the proposed *SHIELD* algorithm.

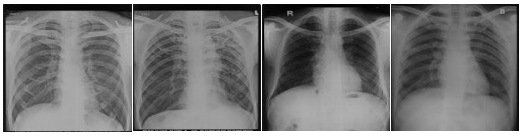

Figure 5: Illustrates the samples of **(A)** Silicosis, **(B)** TB, **(C)** STB and **(D)** Normal.

consistency, we implemented: 1. Systematic task distribution among radiologists, 2. Initial training sessions led by senior experts and 3. Quality control through 20% case review. The dataset encompasses both local and global labels, structured in XML and text formats. Local findings include nodules, consolidation, ground glass opacity, and pleural effusion, while global categories comprise Silicosis, STB, TB, and normal cases. To generate the segmentation masks and findings (local labels), we used Sketchbook [Ske, 2025]and Labellmg [Tzutalin, 2015] software, respectively. The annotation process incorporated several key improvements, including standardized annotation protocols through the Labellmg toolbox, a hierarchical review system to reduce inter-observer variability, and comprehensive documentation of image quality issues. The project faced technical challenges related to variable image quality in digitized radiographs, chemical artifacts and storage-related degradation, and the complexity of multi-label, multi-class annotation requirements. This systematic approach to annotation created a solid foundation for developing deep learning models specifically focused on chest radiograph analysis, with particular emphasis on Silicosis and TB detection. The sample annotation in XML and text, along with the report, is provided in Fig.4.

Evaluation Metrics: To assess the performance of the proposed model for classifying the given CXRs for the four aforementioned classes, we used the Area Under the Receiver Operating Characteristic (AUROC). We report the AUC (one vs rest) values to compare the performance across the proposed approach and baseline models (Table 2). We also report the class-wise accuracy for each class in Table 3.

Implementation Details: The proposed approach, along with the baseline, is implemented using the Pytorch framework [Paszke *et al.*, 2019]. The pretrained models are downloaded from the official website of Pytorch[1]. We used the OpenCV library images to resize the images to 1024, 512, and 256 resolutions. The models are trained on NVIDIA V100-DGXS GPUs for 70 epochs for pretext and downstream tasks using the cross-entropy loss function. We fixed the batch size of 16 across all the tasks, including the segmentation and Adam [Kingma and Ba, 2015] as the optimizer. We used the learning rates of 1e-4, 1e-5, and 1e-4 for segmentation, pretext, and downstream tasks, respectively.

[1]https://pytorch.org/

Approach	Classes	DenseNet121	ResNet18	VGG16	AlexNet
Transfer Learning	STB	0.75	0.79	0.72	0.70
	TB	0.69	0.66	0.63	0.60
	Normal	0.70	0.65	0.68	0.65
	Silicosis	0.72	0.63	0.65	0.61
	Average	0.71	0.68	0.67	0.64
SSL (256)	STB	0.84	0.85	0.80	0.79
	TB	0.82	0.85	0.79	0.81
	Normal	0.86	0.85	0.82	0.82
	Silicosis	0.81	0.83	0.79	0.80
	Average	0.83	0.84	0.80	0.80
SSL (512)	STB	0.87	0.87	0.84	0.81
	TB	0.84	0.87	0.81	0.81
	Normal	0.87	0.87	0.85	0.85
	Silicosis	0.83	0.84	0.82	0.82
	Average	0.85	0.86	0.85	0.61
SSL (1024)	STB	0.89	0.89	0.86	0.83
	TB	0.87	0.88	0.83	0.82
	Normal	0.89	0.88	0.87	0.88
	Silicosis	0.83	0.86	0.83	0.84
	Average	0.85	0.87	0.84	0.84
Proposed	STB	**0.92**			
	TB	**0.91**			
	Normal	**0.93**			
	Silicosis	**0.89**			
	Average	**0.91**			

Table 2: Illustrates comparing the baseline approaches and *SHIELD*. The bold results are the best results obtained with the proposed two-level weighted scores fusion. The evaluation metric used AUC score (one vs rest).

4 Results and Discussion

To assess the performance of the proposed approach, we conducted the experiments on the collected dataset on Silicosis. To train the model, we resize the given CXR images to a resolution of 224 $\times$ 224, which is an ideal and common image resolution to train deep learning models, preserving the image quality and computational complexity, making it a practical choice for many computer vision applications. We compared the proposed approach with the traditional supervised learning based finetuning approach, encompassing the transfer learning technique. The transfer learning (TL) based finetuning is treated as the baseline for our experimentation. With the proposed approach, we first fine-tune the models obtained in the pretext task. The finetuning is performed at all three resolutions, on the samples with compiled dataset for four-class classification. The four classes are Silicosis, STB, TB, and normal. The results are reported using AUC score (one vs rest) and classwise accuracy. The obtained results are higher than the baseline approaches, highlighted in bold in Table 2 and 3. TL uses knowledge from large datasets; however, it may struggle with domain mismatch in medical imaging, while SSL learns directly from unlabeled medical data, often yielding better generalization and robustness on small, domain-specific datasets. Our comprehensive experimental evaluation demonstrates the efficacy of the proposed SSL framework compared to conventional transfer learning methodologies across multiple quantitative metrics and architectural configurations. This section presents an in-depth analysis of the classification performance, architectural comparisons, failure modes, and interpretability aspects.

Classification Performance Metrics: As shown in Table 3, the class-wise accuracy analysis reveals a consistent performance gradient with increasing SSL feature dimensionality. The baseline transfer learning implementations exhibited moderate performance characteristics, with DenseNet121 and

Approach	Classes	DenseNet121	ResNet18	VGG16	AlexNet
Transfer Learning	STB	0.52	0.55	0.53	0.51
	TB	0.61	0.59	0.52	0.50
	Normal	0.65	0.60	0.59	0.53
	Silicosis	0.53	0.58	0.52	0.55
	Average	0.57	0.57	0.54	0.52
SSL (256)	STB	0.54	0.48	0.51	0.61
	TB	0.65	0.64	0.59	0.62
	Normal	0.67	0.64	0.62	0.57
	Silicosis	0.54	0.64	0.60	0.59
	Average	0.60	0.60	0.58	0.59
SSL (512)	STB	0.60	0.57	0.52	0.62
	TB	0.70	0.70	0.62	0.63
	Normal	0.68	0.70	0.64	0.62
	Silicosis	0.62	0.63	0.61	0.58
	Average	0.65	0.65	0.59	0.61
SSL (1024)	STB	0.70	0.64	0.54	0.62
	TB	0.72	0.80	0.67	0.70
	Normal	0.69	0.72	0.65	0.67
	Silicosis	0.66	0.85	0.64	0.60
	Average	0.69	0.70	0.62	0.64
Proposed	STB	**0.65**			
	TB	**0.81**			
	Normal	**0.74**			
	Silicosis	**0.67**			
	Average	**0.72**			

Table 3: Comparing baseline approach with *SHIELD* in terms of class-wise accuracy. The bold values represent the best results obtained with the proposed two-level weighted scores fusion.

ResNet18 achieving comparable mean accuracies of 0.57 across classes, while VGG16 and AlexNet demonstrated marginally reduced performance at 0.54 and 0.52, respectively. The initial SSL-256 configuration showed measurable improvements over the baseline, with mean accuracies ranging from 0.58 to 0.60 across architectural variants, as detailed in Table 1. This performance enhancement became more pronounced with SSL-512, where both DenseNet121 and ResNet18 achieved mean accuracies of 0.65. The most substantial performance gains were observed with SSL-1024, where ResNet18 demonstrated particularly strong classification capabilities for TB (0.80) and Silicosis (0.85). Our proposed fusion methodology further elevated these results, achieving a superior overall mean accuracy of 0.72. As evidenced in Table 1, notable performance peaks were observed in TB detection (0.81), followed by Normal (0.74), Silicosis identification (0.67), and STB (0.65). This represents a significant performance gain of approximately 15% compared to the baseline transfer learning models.

ROC Analysis and AUC Metrics: The Area Under the Curve (AUC) metrics, presented in Table 2, provide additional validation of our methodology's discriminative capabilities. The baseline transfer learning approaches demonstrated moderate performance with AUC scores ranging from 0.64 to 0.71. The SSL variants showed systematic improvements, with SSL-256 achieving mean AUC scores between 0.80 and 0.84 across architectural configurations.

SSL-512 further enhanced performance with mean AUC scores reaching 0.85-0.86 for DenseNet121 and ResNet18. The SSL-1024 variant maintained this elevated performance level, with individual class AUC scores peaking at 0.89 for specific categories. The proposed fusion approach achieved optimal performance with a mean AUC score of 0.91, demonstrating better discriminative capability across all classes (0.92, 0.91, 0.93, and 0.89 for respective classes).

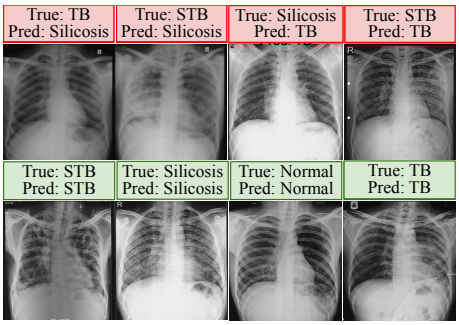

Figure 6: Images of misclassified and correctly classified samples.

Confusion Matrix and Error Analysis: The confusion matrices illustrated in Table. 4 provide crucial insights into classification behaviour and error modes across different methodological approaches. The transfer learning baseline exhibited significant inter-class confusion, particularly between TB and Silicosis, with elevated misclassification rates. The SSL variants demonstrated progressive reduction in these misclassification patterns. The SSL-512 configuration showed marked improvement in correct classifications, particularly for Silicosis (192 correct predictions) and TB (158 correct predictions). The SSL-1024 variant further enhanced this performance profile, achieving 219 correct Silicosis predictions and 180 correct TB predictions. The proposed fusion approach achieved maximal correct classifications across all categories, with particular strength in Silicosis detection (224 correct predictions) and TB identification (182 correct predictions).

Failure Case Analysis: Table 4 shows several failure cases through the confusion matrices. The most prominent challenge involves distinguishing between TB and Silicosis in cases where radiographic patterns show overlapping characteristics. Analysis of false positives indicates that the model occasionally misinterprets diffuse nodular patterns characteristic of Silicosis as TB, particularly in cases with extensive bilateral involvement. As shown in the misclassification examples in Figure 6 (top row), another significant failure mode involves the misclassification of STB cases, where the presence of both silicotic and tuberculous changes creates complex radiographic patterns. The model demonstrated increased uncertainty in cases where the temporal progression of the disease resulted in architectural distortion, making it difficult to distinguish primary pathology from secondary complications.

GradCAM Visualization and Interpretability: The GradCAM [Selvaraju *et al.*, 2017] visualization analysis presented in Figure 7 provides crucial insights into the model's decision-making process and attention mechanisms. For Silicosis cases (top two rows), the activation maps focus on characteristic nodular patterns in upper lung zones, particularly emphasising bilateral symmetrical distribution patterns. In STB cases (middle two rows), the visualization reveals dual activation patterns focusing on both nodular opacities and areas of consolidation. This learning aligns with the complex radiographic manifestations of combined silicosis and TB. TB cases (bottom two rows) show distinct activation patterns emphasizing areas of infiltrates, cavitary lesions, and

(a) Transfer Learning

True/Pred	STB	TB	Nor	Sil
STB	118	45	32	32
TB	40	137	25	23
Nor	20	15	94	15
Sil	55	48	47	169

(b) SSL-256

STB	TB	Nor	Sil
123	42	30	32
35	146	24	20
18	12	97	17
51	45	45	178

(c) SSL-512

STB	TB	Nor	Sil
136	38	25	28
30	158	20	17
15	10	98	21
46	40	41	192

(d) SSL-1024

STB	TB	Nor	Sil
145	35	25	22
25	180	12	8
12	8	104	20
40	25	35	219

(e) Proposed

STB	TB	Nor	Sil
148	35	24	20
20	182	13	10
10	8	107	19
38	22	35	224

Table 4: Confusion matrices comparing the classification performance of different approaches. (Nor:Normal and Sil: Silicosis).

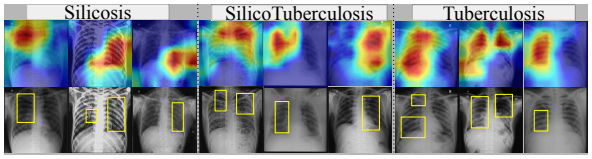

Figure 7: The visualization demonstrates how the proposed deep learning framework identifies key regions in CXRs for Silicosis diagnosis through GradCAM activation mapping. The model effectively focuses on relevant lung areas across all four diagnostic categories, validating its feature detection capabilities.

upper lobe predominance. The model pays particular attention to pulmonary TB's asymmetric distributions and pleural involvement characteristics.

5 Case Study with Real World Impact

Our research implementation directly supports multiple United Nations Sustainable Development Goals while embodying the UN's "Leave No One Behind" principle through technology-enabled healthcare access in marginalized industrial communities. We developed our initial model using a comprehensive dataset from an aspirational district under India's Aspirational Districts Programme (ADP)[2], comprising 3,044 high-resolution CXRs (average size 3567 × 2898) collected over three years. To rigorously validate the model's real-world impact, we conducted extensive testing using a completely independent dataset collected from four geographically dispersed PHCs, strategically chosen to represent diverse healthcare settings distinct from our training region of the collected dataset.

Supporting the UN's "Leave No One Behind" agenda, we developed an intuitive GUI-based diagnostic tool (Fig. 8) specifically designed for and currently being tested by healthcare professionals - the intended end-users without technical expertise in AI or computer systems. This real-world clinical validation involves doctors, nurses, and healthcare workers actively using the system in their daily patient care routines across multiple healthcare centres. The ongoing testing process emphasizes practical utility in clinical settings, with continuous feedback from healthcare professionals shaping the tool's refinement. This user-centric approach ensures the system's seamless integration into existing healthcare workflows while maintaining diagnostic accuracy in resource-constrained environments. The current phase of rigorous clinical testing by actual end-users represents a crucial step toward developing a practically viable solution that can be effectively deployed across diverse healthcare settings, demonstrating how technological innovation can be made accessible and useful for frontline healthcare delivery.

[2]https://vajiramandravi.com/aspirational-district-programme/

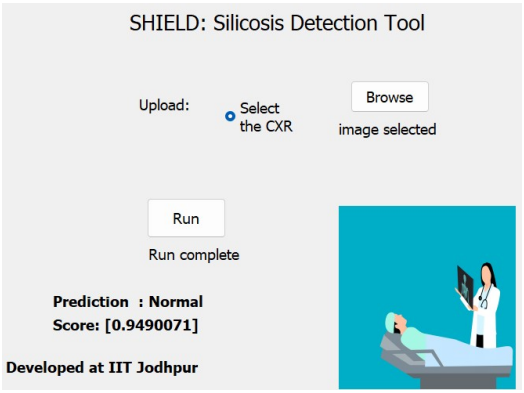

Figure 8: GUI of *SHIELD* for Silicosis diagnosis. The GUI provides the prediction with class-specific scores across four categories: Silicosis, STB, TB, and Normal.

6 Conclusion

We introduce *SHIELD*, a novel self-supervised, multi-resolution jigsaw puzzle mechanism that simultaneously learns features at three distinct scales (1024 × 1024, 512 × 512, and 256 × 256), enabling the model to capture both fine-grained pathological patterns and global anatomical structures in CXRs for diagnosing Silicosis. Unlike conventional single-resolution approaches, the proposed method preserves crucial diagnostic information across scales, vital for detecting diffuse diseases (requiring global context) and localized abnormalities (requiring fine detail) in CXRs. The jigsaw puzzle design incorporates domain-specific constraints based on thoracic anatomy, ensuring that the patch arrangements preserve meaningful anatomical relationships. This anatomically informed approach guides the model to learn clinically relevant features during the self-supervised phase, resulting in more robust and interpretable representations for downstream diagnostic tasks. The proposed two-stage ensemble pipeline leverages the complementary strengths of different spatial resolutions. The first stage processes resolution-specific features through specialized models, while the second stage employs an adaptive weighted score fusion to combine these multi-scale predictions optimally. This hierarchical structure significantly improves the model's ability to differentiate between visually similar conditions, such as silicosis and TB, which present distinct patterns at different spatial scales.

Ethical Statement

The data comprises anonymized CXRs collected from patients under the Institute Review Board's approved guidelines (IIT Jodhpur, Ref No. IEC/IITJ/2022-23/03), with all necessary permissions obtained from PHCs. Patient consent was waived, and data was fully deidentified, ensuring no personal or center-identifying information was collected or released.

Acknowledgments

This work is supported by Srijan: Center of Excellence on GenAI at IIT Jodhpur, India, IndiaAI Mission, and Meta.

References

[Akhter *et al.*, 2023a] Yasmeena Akhter, Rishabh Ranjan, et al. On AI-assisted pneumoconiosis detection from chest x-rays. In *Proceedings of the Thirty-Second IJCAI*, pages 6353–6361, 8 2023. AI for Good - Projects.

[Akhter *et al.*, 2023b] Yasmeena Akhter, Richa Singh, and Mayank Vatsa. AI-based radiodiagnosis using chest x-rays: A review. *Frontiers in Big Data*, 6, 2023.

[Akhter *et al.*, 2025] Yasmeena Akhter, Rishabh Ranjan, Mayank Vatsa, Richa Singh, Santanu Chaudhury, Anjali Agarwal, Shruti Aggarwal, Arjun Kalyanpur, and Anurita Menon. Silicodata: An annotated benchmark CXR dataset for silicosis detection. *Nature Scientific Data*, 2025.

[Alowais *et al.*, 2023] Shuroug A Alowais, Sahar S Alghamdi, et al. Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC medical education*, 23(1):689, 2023.

[Azizi *et al.*, 2021] Shekoofeh Azizi, Basil Mustafa, et al. Big self-supervised models advance medical image classification. In *Proceedings of the IEEE/CVF ICCV*, pages 3478–3488, 2021.

[Barnes *et al.*, 2019] Hayley Barnes, Nicole SL Goh, et al. Silica-associated lung disease: an old-world exposure in modern industries. *Respirology*, 24(12):1165–1175, 2019.

[Cowie, 1994] Robert L Cowie. The epidemiology of tuberculosis in gold miners with silicosis. *American journal of respiratory and critical care medicine*, 150(5):1460–1462, 1994.

[Davies, 1957] CN Davies. Pneumoconiosis, silicosis, and the physics and chemistry of dust. *Annual Review of Medicine*, 8(1):323–348, 1957.

[Devnath *et al.*, 2021] Liton Devnath, Suhuai Luo, et al. Automated detection of pneumoconiosis with multilevel deep features learned from chest x-ray radiographs. *Computers in Biology and Medicine*, 129:104125, 2021.

[Ehrlich *et al.*, 2006] RI Ehrlich, GJ Churchyard, et al. Tuberculosis and silica exposure in south african gold miners. *Occupational and environmental medicine*, 63(3):187–192, 2006.

[Ehrlich *et al.*, 2021] Rodney Ehrlich, Paula Akugizibwe, et al. The association between silica exposure, silicosis and tuberculosis: a systematic review and meta-analysis. *BMC Public Health,*, 2021.

[Eisemann *et al.*, 2025] Nora Eisemann, Stefan Bunk, et al. Nationwide real-world implementation of ai for cancer detection in population-based mammography screening. *Nature Medicine*, pages 1–8, 2025.

[Göndöcs and Dörfler, 2024] Dóra Göndöcs and Viktor Dörfler. AI in medical diagnosis: AI prediction & human judgment. *Artificial Intelligence in Medicine*, 149:102769, 2024.

[Haghighi *et al.*, 2024] Fatemeh Haghighi, Mohammad Reza Hosseinzadeh Taher, et al. Self-supervised learning for medical image analysis: Discriminative, restorative, or adversarial? *Medical Image Analysis*, 94:103086, 2024.

[He *et al.*, 2016] Kaiming He, Xiangyu Zhang, et al. Deep residual learning for image recognition. In *IEEE CVPR*, pages 770–778, 2016.

[Huang *et al.*, 2017] Gao Huang, Zhuang Liu, et al. Densely Connected Convolutional Networks . In *IEEE CVPR*, pages 2261–2269, Los Alamitos, CA, USA, July 2017. IEEE Computer Society.

[Huang *et al.*, 2023a] Shih-Cheng Huang, Anuj Pareek, et al. Self-supervised learning for medical image classification: a systematic review and implementation guidelines. *NPJ Digital Medicine*, 6(1):74, 2023.

[Huang *et al.*, 2023b] Xuezan Huang, Wei Liu, et al. 30-year trends in the disease burden, incidence, and prevention of pneumoconiosis. *China CDC Weekly*, 5(38):856, 2023.

[Jaeger *et al.*, 2014] Stefan Jaeger, S. Candemir, et al. Two public chest x-ray datasets for computer-aided screening of pulmonary diseases. *Quantitative imaging in medicine and surgery*, 4 6:475–7, 2014.

[Kingma and Ba, 2015] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Proceedings at 3rd ICLR, San Diego, CA, USA, May 7-9*, 2015.

[Krizhevsky *et al.*, 2012] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in NeurIPS*, 25:1097–1105, 2012.

[Leung *et al.*, 2012] Chi Chiu Leung, Ignatius Tak Sun Yu, and Weihong Chen. Silicosis. *The Lancet*, 379(9830):2008–2018, 2012.

[Liu and Deng, 2015] Shuying Liu and Weihong Deng. Very deep convolutional neural network based image classification using small training sample size. In *2015 3rd IAPR ACPR*, pages 730–734, 2015.

[Mandrioli *et al.*, 2018] Daniele Mandrioli, Vivi Schlünssen, et al. WHO/ILO work-related burden of disease and injury: protocol for systematic reviews of occupational exposure to dusts and/or fibres and of the effect of occupational exposure to dusts and/or fibres on pneumoconiosis. *Environment international*, 119:174–185, 2018.

[Manna *et al.*, 2024] Siladittya Manna, Saumik Bhattacharya, and Umapada Pal. Self-supervised visual representation learning for medical image analysis: A comprehensive survey. *TMLR*, 2024.

[Mirbabaie *et al.*, 2021] Milad Mirbabaie, Stefan Stieglitz, and Nicholas RJ Frick. Artificial intelligence in disease diagnostics: A critical review and classification on the current state of research guiding future direction. *Health and Technology*, 11(4):693–731, 2021.

[Momtazmanesh *et al.*, 2023] Sara Momtazmanesh, Sahar Saeedi Moghaddam, et al. Global burden of chronic respiratory diseases and risk factors, 1990–2019: an update from the global burden of disease study 2019. *EClinicalMedicine*, 59, 2023.

[Mushtaq *et al.*, 2024] Faisel Mushtaq, Saunak Bhattacharjee, et al. Artificial intelligence for computer aided detection of pneumoconiosis: A succinct review since 1974. *Engineering Applications of Artificial Intelligence*, 133:108516, 2024.

[Neelsen *et al.*, 2022] Sven Neelsen, Patrick Hoang-Vu Eozenou, et al. The 2022 update of the health equity and financial protection indicators database: An overview. Health, Nutrition, and Population (HNP) Discussion Paper, World Bank Group, Washington, D.C., 2022.

[Okumura *et al.*, 2011] Eiichiro Okumura, Ikuo Kawashita, and Takayuki Ishida. Computerized analysis of pneumoconiosis in digital chest radiography: effect of artificial neural network trained with power spectra. *Journal of digital imaging*, 24(6):1126–1132, 2011.

[Okumura *et al.*, 2017] Eiichiro Okumura, Ikuo Kawashita, and Takayuki Ishida. Computerized classification of pneumoconiosis on digital chest radiography artificial neural network with three stages. *Journal of digital imaging*, 30:413–426, 2017.

[Paszke *et al.*, 2019] Adam Paszke, Sam Gross, et al. Pytorch: An imperative style, high-performance deep learning library. In *Advances in NeurIPS, Vancouver, BC, Canada*, pages 8024–8035, 2019.

[Pattichis *et al.*, 2002] Marios S Pattichis, Constantinos S Pattichis, et al. A screening system for the assessment of opacity profusion in chest radiographs of miners with pneumoconiosis. In *Proceedings Fifth IEEE Southwest Symposium on Image Analysis and Interpretation*, pages 130–133. IEEE, 2002.

[Rajpurkar *et al.*, 2017] Pranav Rajpurkar, Jeremy Irvin, Kaylie Zhu, Brandon Yang, Hershel Mehta, Tony Duan, Daisy Ding, Aarti Bagul, Curtis Langlotz, Katie Shpanskaya, et al. Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning. *arXiv preprint arXiv:1711.05225*, 2017.

[Ronneberger *et al.*, 2015] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, volume 9351, pages 234–241. Springer, 2015.

[Selvaraju *et al.*, 2017] Ramprasaath R Selvaraju, Michael Cogswell, et al. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *IEEE ICCV*, pages 618–626, 2017.

[Ske, 2025] Sketchbook — Apps for desktop and mobile devices — sketchbook.com. https://www.sketchbook.com/apps, 2025. [Accessed 25-12-2025].

[Sundararajan *et al.*, 2010] Ramasubramanian Sundararajan, Hao Xu, et al. A multiresolution support vector machine based algorithm for pneumoconiosis detection from chest radiographs. In *2010 IEEE ISBI: From Nano to Macro*, pages 1317–1320. IEEE, 2010.

[Szegedy *et al.*, 2015] Christian Szegedy, Wei Liu, et al. Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–9, 2015.

[Tzutalin, 2015] Tzutalin. LabelImg. Git Code. https://viso.ai/computer-vision/labelimg-for-image-annotation/, 2015.

[Wang *et al.*, 2020] Xiaohua Wang, Juezhao Yu, et al. Potential of deep learning in assessing pneumoconiosis depicted on digital chest radiography. *Occupational and Environmental Medicine*, 77(9):597–602, 2020.

[Wang *et al.*, 2021] Zheng Wang, Qingjun Qian, et al. Deep learning for computer-aided diagnosis of pneumoconiosis. *Research Square*, 2021.

[World Health Organization, 2021] World Health Organization. *WHO global air quality guidelines: particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*. WHO, Geneva, 2021.

[Xu *et al.*, 2010] Horace Xu, Xiaodong Tao, et al. Computer aided detection for pneumoconiosis screening on digital chest radiographs. In *Proceedings of the Third International Workshop on Pulmonary Image Analysis, September 20*, pages 129–138, 2010.

[Yang *et al.*, 2021] Fan Yang, Zhi-Ri Tang, Jing Chen, Min Tang, Shengchun Wang, Wanyin Qi, Chong Yao, Yuanyuan Yu, Yinan Guo, and Zekuan Yu. Pneumoconiosis computer aided diagnosis system based on x-rays and deep learning. *BMC Medical Imaging*, 21(1):189, 2021.

[Yu *et al.*, 2011] Peichun Yu, Hao Xu, et al. An automatic computer-aided detection scheme for pneumoconiosis on digital chest radiographs. *Journal of digital imaging*, 24:382–393, 2011.

[Zhang *et al.*, 2021] Liuzhuo Zhang, Ruichen Rong, et al. A deep learning-based model for screening and staging pneumoconiosis. *Scientific reports*, 11(1):1–7, 2021.

[Zhou *et al.*, 2019] Zongwei Zhou, Vatsal Sodha, et al. Models genesis: Generic autodidactic models for 3d medical image analysis. In *MICCAI*, pages 384–393. Springer, 2019.

[Zhu *et al.*, 2014a] Biyun Zhu, Hui Chen, et al. Support vector machine model for diagnosing pneumoconiosis based on wavelet texture features of digital chest radiographs. *Journal of digital imaging*, 27:90–97, 2014.

[Zhu *et al.*, 2014b] Biyun Zhu, Wei Luo, et al. The development and evaluation of a computerized diagnosis scheme for pneumoconiosis on digital chest radiographs. *Biomedical engineering online*, 13:1–14, 2014.