Deep Learning for Condition Detection in Chest Radiographs: A Performance Comparison of Different Radiograph Views and Handling of Uncertain Labels

Mubashir Ahmad
Department of Computer Science
University of Hertfordshire
Hatfield, UK
email: m.ahmad21@herts.ac.uk

Kheng Lee Koay
Department of Computer Science
University of Hertfordshire
Hatfield, UK
email: k.l.koay@herts.ac.uk

Yi Sun
Department of Computer Science
University of Hertfordshire
Hatfield, UK
email: y.2.sun@herts.ac.uk

Vijay Jayaram
Consultant Radiologist
The Princess Alexandra Hospital
Harlow, UK
email: vijay.jayaram@nhs.net

Ganesh Arunachalam
Consultant in Elderly Care
The Princess Alexandra Hospital
Harlow, UK
email: ganesh.arunachalam@nhs.net

Farshid Amirabdollahian
Department of Computer Science
University of Hertfordshire
Hatfield, UK
email: f.amirabdollahian2@herts.ac.uk

***Abstract*—Chest radiographs are the initial diagnostic modality for lung or chest-related conditions. It is believed that radiologist's availability is a bottleneck impacting patient's safety because of long waiting times. With the arrival of machine learning and especially deep learning, the race for finding Artificial Intelligence (AI) based approaches that allow for the highest accuracy in detecting abnormalities on chest radiographs is at its peak. Classification of radiographs as normal or abnormal is based on the training and expertise of the reporting radiologist. The increase in the number of chest radiographs over a period of time and the lack of sufficient radiologists in the UK and worldwide have had an impact on the number of chest radiographs assessed and reported in a given time frame. Substantial work is dedicated to machine learning for classifying normal and abnormal radiographs based on a single pathology. The success of deep learning techniques in binary radiograph classification urges the medical imaging community to apply it to multi-label radiographs. Deep learning techniques often require huge datasets to train its underlying model. Recently, the availability of large multi-label datasets has ignited new efforts to overcome this challenging task. This work presents Convolutional Neural Networks (CNNs) based models trained on publically available CheXpert multi-label data. Based on common pathologies seen on chest radiographs and their clinical significance, we have chosen pathologies such as Pulmonary oedema, Cardiomegaly, Atelectasis, Consolidation and Pleural effusion. We trained our models using different projections such as Anteroposterior (AP), Posteroanterior (PA), and lateral and compared the performance of our models for each projection. Our results demonstrate that the model for the AP projection outperforms the remaining models with an average AUC of 0.85. Furthermore, we use the samples with uncertain labels in CheXpert dataset and improve the model performance by removing the uncertainty using Gaussian Mixture Models (GMM). The results show improvement in all three views with AUCs ranging from 0.91 for AP, 0.75 for PA and 0.85 on the lateral view.**

Keywords– ***Chest radiograph; Deep learning; Multi-label classification; Uncertain labels***

I. Introduction

Respiratory diseases are one of the leading causes of death in the United Kingdom. According to a survey by Conor Stewart [1], prior to COVID-19, the mortality rate from respiratory diseases in the United Kingdom in 2020 was 130 per 100,000 male population and 89 per 100,000 female population. Chest radiographs are the most utilised diagnostic modality for lung or chest-related conditions. However, it requires an experienced radiologist to accurately analyse radiographs to detect chest-related conditions, such as Pulmonary oedema, Cardiomegaly, Atelectasis, Lung cancer, and Consolidation, besides other less common pathologies. A report in 2020 by the Royal College of radiologists [2] highlights the national shortage of radiologists resulting in reporting backlogs which can adversely impact patient care. A further predicted shortfall in radiologist numbers by 44% in 2025 will have a greater impact on reporting backlogs. Expenditure on outsourcing imaging examinations for reporting has increased by 58% in the last few years. In addition to the scarcity of radiologists, there is a problem with diagnostic errors in radiology reports. According to [3], worldwide, annually, at least 40 million out of 1 billion radiology reports contain errors. Chest radiographs are the most used diagnostic procedure for respiratory and cardiovascular diseases - errors in diagnosis and delays in reporting contribute to adverse outcomes for patients. In order to decrease the workload for existing and future radiologists, scientists have been working on automatic radiographic interpretation systems. Recently, Deep Learning especially Convolutional Neural Networks (CNNs) has boosted research in computer vision, especially in medical imaging and has demonstrated promising results in the detection of pathologies on chest radiographs. However, the interpretation of chest radiographs can be challenging. In order to provide good results, deep learning requires a large number of data samples for training. The presence of multiple conditions in one radiograph makes it difficult for the model to generalise as there can be an overlap in the imaging findings of two different chest pathologies, e.g., pulmonary oedema and infectious pathologies. Moreover, the presence of uncertain labels in a dataset further increases the difficulty. Samples with uncertain labels can cause difficulties in training the machine learning algorithm. Besides demonstrating the potential to improve diagnostic accuracy, deep learning models can also improve the reporting workflow by prioritising abnormal radiographs over normal ones. Training the model on different radiographic projections has the potential to improve the diagnostic accuracy of the model, particularly when dealing with a suboptimal radiograph in a critically ill patient.

This study trains a state-of-the-art CNN-based deep neural network DenseNet121 [4] on a large chest radiograph dataset, CheXpert [5][6]. Figure 1 shows a few radiographs from the CheXpert dataset. In addition, we trained separate models for different views Anteroposterior (AP), Posteroanterior (PA), and lateral and evaluated their performance. Furthermore, the study addresses uncertainty present in the data by relabelling uncertain samples using a Gaussian Mixture Model (GMM) [7] and including them in the training data. The performance is then compared before and after reducing the uncertainty. The following Section offers a review of the state of the art using deep learning for interpreting chest radiographs, identifying multiple conditions in a single radiograph, and addressing uncertain labels in data. Section III presents our approach, which involves utilizing various data and model-based methods, correcting uncertain labels, and addressing each radiographic view individually. In Section IV, we conduct comparative experiments on the CheXpert dataset and provide the results. Section V examines the main insights gained from the results. Lastly, in Section VI, we present the conclusion of the paper.

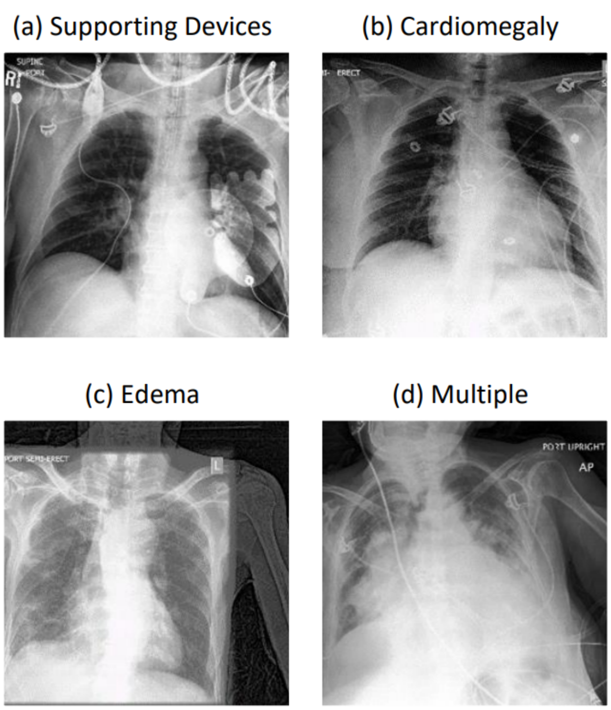

Figure 1: Images from CheXpert dataset. Each image has 14 labels corresponding to each condition. Mentioned conditions are positive labels for each of the four images.

II. Related Work

The availability of large labelled datasets [5][6][8] of chest radiographs has led many researchers to use deep learning for chest radiograph interpretation. Most recent work in this area has focused on CNN-based models and applied various techniques such as transfer learning, feature extraction, and region of interest analysis to improve the detection of abnormalities [9–12]. In one study, a 121-layered neural network trained on the CheXnet frontal view dataset, outperformed average radiologists in detecting pneumonia [9]. Additionally, data augmentation has been used to tackle the issue of insufficient data in new challenges such as COVID-19, as seen in [13] where a CNN model was trained to classify chest radiographs of COVID-19 patients. Many studies have focused on specific conditions such as pneumonia, COVID-19 and oedema [9][13][14]. In addition to the above, the power of deep learning allows for the detection of multiple conditions in a single radiograph [15][16]. Multiple-label detection on chest radiographs is much more challenging as compared to a single label. The overlapping and vanishing of features can hinder the model performance in a multi-label setting [17]. The hierarchical dependencies present between conditions are exploited in [15] by using a conditional training approach. This is achieved by training a deep neural network twice, first on data with only positive parent-level conditions, followed by training on the entire dataset. As CNN is very good at extracting prominent features, [18] suppresses the irrelevant features by assigning them smaller weights and enhancing the important features with higher weights to detect multiple conditions in chest radiographs. Different abnormal conditions appear on radiographs in different anatomical areas, such as a Pleural effusion, which can be identified by looking at the lower left and right corners of the lungs. Localising the correct anatomical region in [19] enables the model to learn the better relationship between different structures in the radiograph.

This paper examines the use of three different radiograph views (AP, PA, lateral) separately. In the first phase, we train a DenseNet121 model for each view, using techniques such as transfer learning, template matching, and augmentation to improve performance. We have repeated these experiments ten times to ensure generalisable results. In the second phase, we use a semi-supervised approach with GMM to label uncertain samples, which is an improvement over previous methods [5][15] that assigned positive labels to all uncertain samples or a random float between 0.55 to 0.85. We then include these relabelled uncertain samples in the training data and repeat all experiments. Our approach of individually analysing each view shows promising results and effectively reduces uncertainty in the data.

III. Methodology

In this section, we outline the method utilised to classify radiographs with multiple labels. We begin by introducing

 ISBN: 978-1-68558-078-0

DenseNet121 and CNN briefly along with the applied model training procedure in subsection A. Following that, we delve into explaining multi-scale template matching for data quality improvement and transfer learning techniques we employed to enhance the model's performance in subsection B. Furthermore, in subsection C, we expound upon the data augmentation methods we adopted to enhance data diversity and control model overfitting. Finally, we describe how we utilize GMM to eliminate uncertainty in the data in subsection D.

A. DenseNet121 and Model Training

In this paper, we chose DenseNet121 architecture as our base CNN model because of its popularity in computer vision, especially in medical imaging [15][20][21]. DenseNet121 utilizes convolutional neural networks (CNN) and comprises 121 layers. The layers in the network are connected in such a way that each layer receives inputs from all the previous layers; this helps the model retain important features recognized in the earlier layers [4]. CNN is a deep learning algorithm that performs a convolution operation on images to extract features. We also used pooling and dropout layers to prevent the model from overfitting. Finally, we trained the model using a batch size of 32, Adam optimizer, and used the Area under the Receiver Operating Characteristic Curve (AUC) as the evaluation metric, as in [15].

B. Multi-scale Template Matching and Transfer Learning

In order to improve the data quality, we employed a multi-scale template matching technique to eliminate unnecessary regions from both the training and testing images [15][22]. To achieve this, we picked a high-quality template image from each view, trimmed and scaled it to 224×224 pixels, and removed unnecessary areas such as the shoulders, neck, and pelvic sections. We then matched the template image at a scale range (empirically chosen) of (0.7 to 1.3) with the entire data and extracted the best-matched 224×224 section of the image, thus enabling the model to concentrate on the thoracic area and avoiding any confusion from other regions.

Furthermore, to help the model train fast and accurately we applied transfer learning. This is a powerful method to improve the accuracy of deep learning models. The idea is to leverage the knowledge gained by training on a generic data set, such as ImageNet [23], to gain knowledge of general image features (vertical and horizontal edges). The last layers of the model are then replaced and fine-tuned the model with data specific to the task at hand, such as CheXpert [5] in this instance. As the data in ImageNet is very different from the one in CheXpert, so instead of fine-tuning just the last layers of the model we fine-tuned all layers[24]. This makes all layers specialised to radiographs while getting help from general image features. Through experimentation, we have compared the performance of models with and without transfer learning. The detail is in section IV of this paper.

C. Data Augmentation

In the medical image classification domain, insufficient data has always been a problem. Chest radiograph data is no different. Although we have very large chest radiograph data sets available [5][6][8], these are still not enough for a deep-learning model to generalise ideally. This is because the same pathology can manifest differently on a radiograph depending on the patient's age, gender, lifestyle and stage of the disease, besides the radiographic projection and other technical factors. Multiple pathologies on a single radiograph make feature extraction more difficult. That is the reason, deep learning algorithms require a very large number of samples with the same pathology to capture sufficient important features. With data augmentation, we artificially enhance the size of the data set and add diversity to it. Various techniques can be used to modify the image, such as resizing and zooming. In order to improve feature extraction, we employed data augmentation on our data set which includes setting brightness randomly between 30% and 100%, randomly rotating the image by $7\pm$ degrees, applying a random shear range of $0.2\pm$, zooming the image by 0.2, adding random noise to the images, and finally flipping the images horizontally [25]. These six augmentation techniques were chosen empirically. We apply all six data augmentation techniques on all training images and send them to the model along with the original image for training. The results of the experiment reveal that applying image augmentation significantly improves performance.

D. Relabelling Uncertain labels with GMM

In the CheXpert dataset, almost 30% of the samples have uncertain labels, which means the condition may or may not be present in the radiograph. As this is a multi-label problem, the presence of one condition can impact the appearance of another coexisting condition on the radiograph. Instead of discarding the 30% of the data with uncertain samples in CheXpert, or assigning all positive/negative labels, we removed the uncertainty and relabelled the samples and include them in the training process. To do that, we chose GMM because of its ability to tackle a mixture of multiple data distributions. It is a probabilistic model that creates multiple clusters using an Expectation Maximization (EM) algorithm and updates the estimator parameters during the process [7]. We trained a GMM model for each of the five conditions separately using only certain samples. GMM operates by assigning each sample to the cluster with the distribution that is closest in terms of parameters. It created 100 clusters for Consolidation, 500 for Pleural Effusion, 200 for Cardiomegaly, 300 for Atelectasis and 350 for Edema. This results in many clusters, with multiple clusters of each class. Once the estimator was fully converged, we used it to get predictions for the uncertain samples. We conducted experiments excluding uncertain samples and then including them after relabelling in the model training. The results indicate that this approach leads to some performance improvement. Further detail is in section IV of this paper.

IV. Experimentation and Results

The dataset has 223,414 chest radiographs of 65,240 patients collected between October 2002 and July 2017. Each

 ISBN: 978-1-68558-078-0

image has 14 labels corresponding to a medical condition. In this study, we chose five clinically important conditions Pulmonary oedema, Consolidation, Cardiomegaly, Atelectasis, and Pleural Effusion [5]. To make the performance comparison between different radiograph projections, we created a separate model for each projection. Also, to ensure fair performance comparison, we trained the models on an equal number of samples (29,421 images per view). We did not use the CheXpert validation set during training and instead used it to test the models after training. The experiments were conducted in two phases. In the first phase, we excluded samples with uncertainty from the model training process, further explained in subsection A. In the second phase of experiments, we incorporated 22,219 relabelled uncertain samples for each view in the training process. Through this method, we were able to observe the impact of eliminating data uncertainty on the performance of the models.

A. Experiments Excluding Uncertain Samples

In this paper, we used DenseNet121 as the main network and trained five different models for each radiograph view. These models include DenseNet121, DenseNet121 with multi-scale template-matched (TM) data, DenseNet121 with transfer learning (TL), DenseNet121 with data augmentation (AUG), and a combination of template matching, transfer learning, and augmentation. The AUC is used as the evaluation metric in all of the experiments done in this paper. Table I shows the results on the Anteroposterior view. We

TABLE I: AUC SCORES FOR VARIOUS DEEP LEARNING METHODS WITH DENSENET121 ON AP, WITHOUT UNCERTAIN SAMPLES. THE VALUES IN BOLD SHOW THE BEST RESULTS OF EACH MODEL.

Anteroposterior					
Exp	**DN121**	**DN121_TM**	**DN121_TL**	**DN121_AUG**	**DN121_TM_TL_AUG**
1	**0.85**	0.79	0.76	**0.87**	0.84
2	0.79	0.74	0.81	0.76	0.85
3	0.76	0.75	0.86	0.79	0.82
4	0.78	0.81	0.8	0.8	0.83
5	0.76	**0.81**	0.84	0.84	0.83
6	0.76	0.72	**0.9**	0.86	0.83
7	0.8	0.79	0.67	0.8	0.88
8	0.82	0.63	0.8	0.84	0.85
9	0.81	0.79	0.82	0.82	0.86
10	0.74	0.78	0.81	0.8	**0.88**
Avg	0.79	0.76	0.81	0.82	0.85
Std	0.03	0.05	0.06	0.03	0.02

repeat each experiment ten times with different sample sets randomly chosen from CheXpert. We can see how different techniques applied with DenseNet121 performed better, especially transfer learning and data augmentation. The best-performing model is "$DN121\_TM\_TL\_AUG$" with statistical significance observed using Analysis of Variance (ANOVA) where $[F(4,45) = 5.504, p = 0.001]$. When repeating the same experiment for the other two views, PA and Lateral, a similar statistical significance is observed in favour of this combined model. The performance of this model, "$DN121\_TM\_TL\_AUG$", is then compared across the three views. Table II shows the results from this comparison. ANOVA results indicate statistically significant differences between the three views where $[F(2,27) = 64.677, p < 0.001]$. Summary statistics indicate that AP and Lateral views performed better than PA. AP performed more consistently with a standard deviation of 0.02 as compared to 0.04 for Lateral.

TABLE II: RESULTS OF EXPERIMENTS USING OUR BEST-PERFORMING MODEL ON AP, PA AND LATERAL VIEWS WITHOUT UNCERTAIN SAMPLES.

Exp	**Anteroposterior(AP)**	**Posteroanterior(PA)**	**Lateral**
1	0.84	0.69	0.79
2	0.85	0.73	**0.9**
3	0.82	0.74	0.84
4	0.83	0.74	0.78
5	0.83	**0.74**	0.87
6	0.83	0.73	0.82
7	0.88	0.73	0.84
8	0.85	0.69	0.86
9	0.86	0.7	0.82
10	**0.88**	0.72	0.81
Avg	0.85	0.72	0.83
Std	0.02	0.02	0.04

B. Experiments Including Uncertain Samples

In the second phase of experiments, we included the relabelled uncertain samples to the training data of the corresponding data set. We reran the whole series of experiments as in the first phase. Table III shows the results of the experiments for the best model of each view. ANOVA results indicate statistical significance for the difference between the three views where $[F(2,27) = 19.823, p < 0.001]$. Interestingly, on average AP is still ahead but if we look at the minimum and maximum AUC, it is 0.65 and 0.91, respectively, also represented by the larger standard deviation for AP.

TABLE III: RESULTS OF EXPERIMENTS USING OUR BEST-PERFORMING MODEL ON AP, PA AND LATERAL VIEWS WITH UNCERTAIN SAMPLES.

Exp	**Anteroposterior(AP)**	**Posteroanterior(PA)**	**Lateral**
1	0.87	0.72	0.78
2	0.65	0.73	**0.85**
3	0.87	0.73	0.84
4	0.89	0.75	0.82
5	**0.91**	0.73	0.77
6	0.9	**0.75**	0.82
7	0.85	0.72	0.83
8	0.88	0.72	0.81
9	0.89	0.73	0.83
10	0.88	0.73	0.83
Avg	0.86	0.73	0.82
Std	0.08	0.01	0.03

Figure 2 shows a better view of gradual improvement in the performance of AP models. The blue boxes indicate the performance of models before including uncertain samples and the orange boxes indicate the performance of models after including uncertain samples which were relabelled using GMM. Performing a Univariate Analysis of Variance for AP identified statistical significance at model level ($p < 0.001$) and also pre-post level ($p = 0.024$), but there was no significance for interaction effects for model and pre-post indicating that we can interpret and rely on significant differences observed at model and pre-post levels.

Overall, this indicates that our GMM approach to reduce uncertain labels contributes to betterment of the classification task.

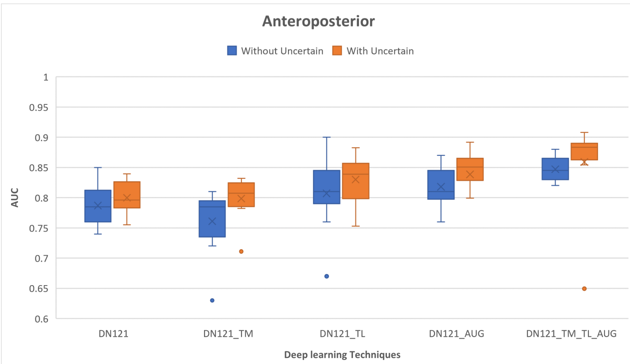

Figure 2: Performance comparison of AP models before and after including uncertain samples.

V. Discussion

By training a state-of-the-art deep learning algorithm on different views of chest radiographs, we identified that AP and Lateral views are comparably better than PA on DenseNet121 and its derivations using some of the state-of-the-art CNN backbone architectures. The process allowed us to identify a model with best performance for the AP and Lateral views. This is an interesting fact, keeping in view that PA radiographs are more commonly performed in the outpatient setting, while AP radiographs are performed in patients who are ill or unstable and therefore unable to cooperate for a PA view. Identifying a condition by a view-specialised model can be more reliable than a general model trained on all types of views. As the frontal and lateral views of the chest look different, important features of a frontal image can manifest differently on a lateral view resulting in uncertainty for interpretation both by the radiologist and the machine learning algorithm.

Machine learning techniques such as GMM can help resolve uncertain labels which can be utilised in the training process as shown in this work. Although the improvement after removing the uncertainty is not significant in terms of AUC increases, this approach shows the potential for further exploration and optimisation. This work has some limitations, for example the data set used contains radiographs from only one hospital. CNN is a data-hungry algorithm, more images are required for each projection to show significant improvements. A further point for investigation is to assess the model performance across different data sets, ideally obtained from different geographical areas.

In the presented work, we applied multiple techniques to improve the data quality and the model's ability. For transfer learning, we utilised a DenseNet121 model pre-trained on ImageNet data. Although the ImageNet dataset differs significantly from CheXpert, it does provide some assistance to the model, but this assistance does not result in any significant increase in AUC. A point for future investigation is to either fine-tuning the model with huge radiograph data or to pre-train the model on a different chest radiograph data and fine-tuning it on CheXpert, this will help the model to learn some basic radiograph features in the pre-training stage. Data augmentation has resulted in significantly improving the model performance. We believe that carefully engineered augmentation techniques can enhance the detection accuracy of radiographs. Counter-intuitively, the multiscale template matching approach did not provide significant benefits and sometimes even resulted in a decrease in performance. Further investigation is necessary to identify the cause of this observation.

The results of our experiments show the potential to improve diagnostic accuracy for chest radiographs and also classify radiographs in a reporting worklist as normal or abnormal thereby prioritising abnormal radiographs for more urgent reporting.

VI. Conclusion

This paper presented a performance comparison of five CNN-based deep learning models trained on different views of chest radiographs. Our results indicated the final derivation of the model utilising a combination of template matching, transfer learning, and augmentation provided the highest average AUC of 0.88. This observation led to choosing the final model as a tool to compare and contrast between different views. Our contrasting led to ordering AP, Lateral and PA views with a decreasing AUC from 0.85 to 0.83 and 0.72, respectively. Improving the model by labelling uncertain samples led to an increase in AUC, by a factor of 0.01, for each of these views. Our results indicated that it is possible to detect and label radiographs in multi-condition images, with anterio-posterior and lateral views outperforming the posterio-anterior view. We also highlighted that our approach to uncertainty reduction can have a positive impact on AUC improvement, indicating better detection accuracy. We now embark on evaluating if there are conditions where different views have a vested advantage in detection, using the above CNN model. We also plan co-design studies with radiologists in our partner hospital, to identify barriers to the acceptability of such models, but also ways to integrate such approaches into the clinical workflow.

References

[1] C. Stewart, "Respiratory disease in the united kingdom (uk) - statistics and facts." https://www.statista.com/topics/5908/respiratory-disease-in-the-uk/ [Retrieved: March, 2023].

[2] RCR, "Clinical radiology uk workforce census 2020 report." https://www.rcr.ac.uk/publication/clinical-radiology-uk-workforce-census-2020-report [Retrieved: March, 2023].

[3] M. A. Bruno, E. A. Walker, and H. H. Abujudeh, “Understanding and confronting our mistakes: the epidemiology of error in radiology and strategies for error reduction,” *Radiographics*, vol. 35, no. 6, pp. 1668–1676, 2015.

[4] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.

[5] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haghgoo, R. Ball, K. Shpanskaya, *et al.*, “Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, pp. 590–597, 2019.

[6] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, “Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2097–2106, 2017.

[7] T. K. Moon, “The expectation-maximization algorithm,” *IEEE Signal processing magazine*, vol. 13, no. 6, pp. 47–60, 1996.

[8] A. E. Johnson, T. J. Pollard, S. J. Berkowitz, N. R. Greenbaum, M. P. Lungren, C.-y. Deng, R. G. Mark, and S. Horng, “Mimic-cxr, a de-identified publicly available database of chest radiographs with free-text reports,” *Scientific data*, vol. 6, no. 1, p. 317, 2019.

[9] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya, *et al.*, “Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning,” *arXiv preprint arXiv:1711.05225*, 2017.

[10] H. Sharma, J. S. Jain, P. Bansal, and S. Gupta, “Feature extraction and classification of chest x-ray images using cnn to detect pneumonia,” in *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, pp. 227–231, IEEE, 2020.

[11] M. Heidari, S. Mirniaharikandehei, A. Z. Khuzani, G. Danala, Y. Qiu, and B. Zheng, “Improving the performance of cnn to predict the likelihood of covid-19 using chest x-ray images with preprocessing algorithms,” *International journal of medical informatics*, vol. 144, p. 104284, 2020.

[12] T. Rahman, M. E. Chowdhury, A. Khandakar, K. R. Islam, K. F. Islam, Z. B. Mahbub, M. A. Kadir, and S. Kashem, “Transfer learning with deep convolutional neural network (cnn) for pneumonia detection using chest x-ray,” *Applied Sciences*, vol. 10, no. 9, p. 3233, 2020.

[13] A. A. Reshi, F. Rustam, A. Mehmood, A. Alhossan, Z. Alrabiah, A. Ahmad, H. Alsuwailem, and G. S. Choi, “An efficient cnn model for covid-19 disease detection based on x-ray image classification,” *Complexity*, vol. 2021, pp. 1–12, 2021.

[14] C. Hayat, “Densenet-cnn architectural model for detection of abnormality in acute pulmonary edema,” *Khazanah Informatika: Jurnal Ilmu Komputer dan Informatika*, vol. 7, no. 2, pp. 73–79, 2021.

[15] H. H. Pham, T. T. Le, D. Q. Tran, D. T. Ngo, and H. Q. Nguyen, “Interpreting chest x-rays via cnns that exploit hierarchical disease dependencies and uncertainty labels,” *Neurocomputing*, vol. 437, pp. 186–194, 2021.

[16] I. M. Baltruschat, H. Nickisch, M. Grass, T. Knopp, and A. Saalbach, “Comparison of deep learning approaches for multi-label chest x-ray classification,” *Scientific reports*, vol. 9, no. 1, pp. 1–10, 2019.

[17] S. Albahli, H. T. Rauf, A. Algosaibi, and V. E. Balas, “Ai-driven deep cnn approach for multi-label pathology classification using chest x-rays,” *PeerJ Computer Science*, vol. 7, p. e495, 2021.

[18] Q. Guan and Y. Huang, “Multi-label chest x-ray image classification via category-wise residual attention learning,” *Pattern Recognition Letters*, vol. 130, pp. 259–266, 2020.

[19] N. N. Agu, J. T. Wu, H. Chao, I. Lourentzou, A. Sharma, M. Moradi, P. Yan, and J. Hendler, “Anaxnet: anatomy aware multi-label finding classification in chest x-ray,” in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24*, pp. 804–813, Springer, 2021.

[20] O. Gozes and H. Greenspan, “Deep feature learning from a hospital-scale chest x-ray dataset with application to tb detection on a small-scale dataset,” in *2019 41st annual international conference of the ieee engineering in medicine and biology society (embc)*, pp. 4076–4079, IEEE, 2019.

[21] J. A. Dunnmon, D. Yi, C. P. Langlotz, C. Ré, D. L. Rubin, and M. P. Lungren, “Assessment of convolutional neural networks for automated classification of chest radiographs,” *Radiology*, vol. 290, no. 2, pp. 537–544, 2019.

[22] R. Brunelli, *Template matching techniques in computer vision: theory and practice*. John Wiley & Sons, 2009.

[23] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255, Ieee, 2009.

[24] Z. Zhou, J. Shin, L. Zhang, S. Gurudu, M. Gotway, and J. Liang, “Fine-tuning convolutional neural networks for biomedical image analysis: actively and incrementally,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7340–7351, 2017.

[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.