

The exigency of an automated model to classify thoracic roentgenograms and provide an accurate diagnosis in the context of patient prodromes and clinical history

Rohan M. Bhansali

Loudoun Academy of Science

30 May 2019

Thoracic radiographs are the most commonly utilised medical diagnostic tool, with over two billion performed annually (Raoof, et al., 2012). However, there is a global shortage of radiologists to analyse these X-rays, as exemplified by the nearly two thirds of the world's population that lacks radiologists. This problem is further exacerbated in poor countries such as Rwanda, where eleven radiologists care for twelve million inhabitants, and Liberia, where, despite a population of four million, there are merely two practising radiologists (Johnson, et al., 2019).

The cardiopulmonary diseases that are typically detected through these radiographs tend to be among the most lethal; they include pneumonia, a contagion that hospitalises over a million Americans annually, of which approximately fifty thousand expire ("Pneumonia Can Be Prevented-Vaccines Can Help | CDC", n.d.); tuberculosis, which currently afflicts one fourth of the world's population and kills an annual average of 1.3 million people worldwide ("Data & Statistics | TB | CDC", n.d.); and lung carcinoma, the deadliest cancer for both men and women, with over one hundred and fifty thousand annual deaths attributed to the disease ("Lung Cancer Is the Biggest Cancer Killer in Both Men and Women" Infographic | CDC", n.d.).

The frequency of chest X-rays and the deadly nature of the diseases associated with them make their accurate diagnosis imperative. However, radiologists, though professionally trained, are subject to human limitations that include fatigue, inattentiveness and bias. Consequently, a model with automated diagnostic capabilities would have enormous consequences; for example, in areas with a deficiency of radiologists, the model could essentially replace the radiologists and provide a diagnosis of patients' X-rays with symptoms and history taken into consideration, just as a radiologist would. A model with this capability would also be extremely versatile as it could be implemented in areas where there are not insufficiencies of radiologists; in these areas, the model could act as a confirmation to the radiologists and could also expedite the diagnostic process and

reduce the costs associated with it. However, to fully simulate the clinical process of diagnosis, a model must meet several criteria.

First, the model must differentiate between anteroposterior (AP) and posteroanterior (PA). Although AP and PA X-rays are both frontal radiographs, they are fundamentally different in terms of the method by which they are performed and the resultant radiograph. PA X-rays are taken from the back to the front whereas AP X-rays are taken from the front to the back. AP X-rays are generally not preferred except in scenarios in which the patient is too weak or is unable to assume an erect position (Puddy & Hill, 2007). AP X-rays are much more difficult to read as radiologists must make several adjustments to account for the differing view. For example, AP X-rays tend to return the appearance of mild cardiomegaly (enlargement of the heart) because the X-rays diverge as they pass through the mediastinum, resulting in an overall magnification of the anterior structures of the thorax, among which is the heart.

Second, the model must take into account the lateral X-rays as they are significant for at least 15% of diagnoses and often reveal information that a frontal X-ray does not (Ittyachen, Vijayan, & Isac, 2017). Lateral X-rays are especially useful in situations where the frontal X-ray is AP; due to their aforementioned difficulty in interpretation, the lateral view often provides clarification and further detail.

Third, the model would need to also account for patient information, including symptoms and medical history. The importance of these factors in the diagnosis can be highlighted by the example of a patient's chest X-ray showing signs of congested lung vasculature. The radiologist analysing the roentgenogram could potentially diagnose it as an acute illness like multifocal pneumonia; however, if the radiologist knew that the patient has a heart condition and is exhibiting shortness of breath, it would be much more likely that the patient is affected by pulmonary edema (Bhansali,

personal communication, March 29, 2019). However, without the knowledge of the patient and the patient's symptoms and history, the radiologist would have provided a misdiagnosis which ultimately could have led to the expiration of the patient.

Several datasets have been released to further the development of machine learning in thoracic radiograph diagnosis. One of the earliest and most significant datasets was the ChestX-ray14 dataset, a large set of thoracic radiographs released by the National Institute of Health. The dataset, which was, at the time of its initial publication, the largest dataset of chest X-rays, contained 112,210 X-ray images in DICOM format from 30,805 patients (Wang, et al., 2017). The dataset was especially notable for the role it played in the development CheXNeXt, a deep learning algorithm designed by researchers belonging to Stanford University's Machine Learning Group. CheXNeXt is a 121-layer convolutional neural network that was trained and validated using the aforementioned ChestX-ray14 dataset (Rajpurkar, et al., 2018). The model took a frontal X-ray as an input and outputted a vector of disease probabilities and a heat map of where the findings of the radiograph were localised.

Figure 1: A visual representation of CheXNeXt, with the inputted X-ray and outputted heat map and disease probability



The model was tested with a set of 420 chest X-rays, which it diagnosed in 90 seconds; conversely, the four board-certified radiologists against whom CheXNeXt was being compared required approximately four hours. Although CheXNeXt was remarkable in terms of the accuracy it achieved, it was limited in that it did not differentiate between AP and PA X-rays, account for lateral chest X-rays or consider patient information. However, despite its shortcomings, CheXNeXt was revolutionary as it was the first model to conclusively provide evidence for the potential of such a model matching, and occasionally surpassing) the accuracy of radiologists.

Table 1: A comparison of the AUC (area under the receiver operating characteristic curve) achieved by radiologists and CheXNeXt for each pathology

Pathology	Radiologists	CheXNeXt	Advantage
Atelectasis	0.808	0.862	CheXNeXt
Cardiomegaly	0.888	0.831	Radiologists
Consolidation	0.841	0.893	No significant difference
Edema	0.910	0.924	No significant difference
Effusion	0.900	0.901	No significant difference
Emphysema	0.911	0.704	Radiologists
Fibrosis	0.897	0.806	No significant difference
Hernia	0.985	0.851	Radiologists
Infiltration	0.734	0.886	No significant difference
Mass	0.886	0.909	No significant difference
Nodule	0.899	0.894	No significant difference
Pleural thickening	0.779	0.798	No significant difference
Pneumonia	0.823	0.851	No significant difference
Pneumothorax	0.940	0.944	No significant difference

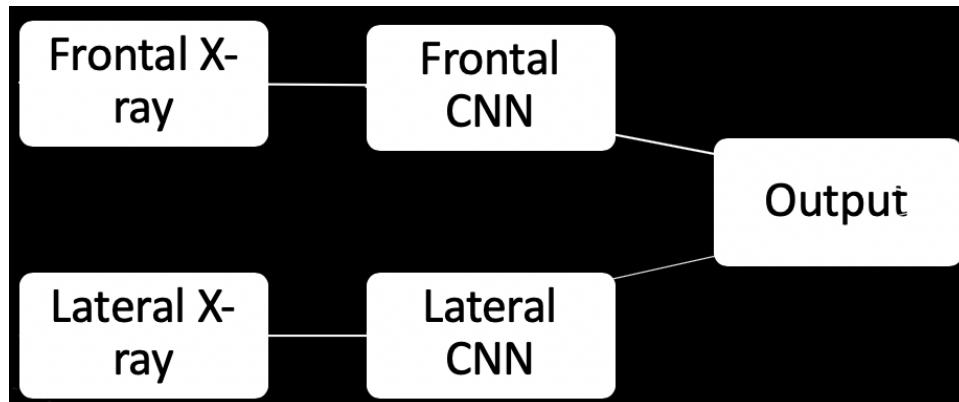
Another prominent dataset is MIMIC-CXR, the largest collection of thoracic radiographs released to date (Johnson, et al., 2019). The dataset contains 371,920 thoracic radiographs with positive and negative labels for the following diseases/findings: no finding¹, enlarged cardiomeastinum,

¹ The “no finding” label indicates the absence of any of the diseases associated with the other labels

cardiomegaly, airspace opacity, lung lesion, pulmonary edema, consolidation, pneumonia, atelectasis, pneumothorax, pleural effusion, pleural other, fracture and support devices. The DICOM-formatted images were obtained from 227,943 radiologic imaging studies conducted at the Beth Israel Deaconess Medical Centre and were de-identified using an algorithm that removed dates and potential patient identifiers. Subsequently, the images were labelled with information from their corresponding radiology report using the CheXpert labeller developed by researchers at Stanford University. The dataset has been published for researcher use and is intended to be fully disseminated in the near future. As a prerequisite to gaining access to the data, completing a course in human ethics is mandatory; additionally, the researchers must agree to citing the data in any publication that makes use of it.

A recently published model developed by researchers at Philips Research Institute was able to build upon the research done by Rajpurkar, et al.; in their research, three separate networks were trained for PA, AP and lateral X-ray images (Rubin, et al., 2018). These networks were then paired together for two different models. One model was composed of a PA and lateral network, while the other model was composed of an AP and lateral network. The model operated by accepting two different inputs: a frontal X-ray and a lateral X-ray. Subsequently, the images were independently analysed through separate networks with the final output being a fusion of the outputs of the individual networks. Each network was designed based on the DenseNet-121 architecture. A sigmoid operation was applied to each of the fourteen outputs and the networks were trained using a binary cross-entropy loss function.

Figure 2: A visual representation of the dual convolutional neural networks



While this model was successful in implementing lateral X-rays, something that prior research had been unable to accomplish, it still did not account for patient information, which would be the focus of future research in this field (Rubin, et al., 2018).

Models of these capabilities would have immense benefits in health-care environments as they would drastically reduce the number of diagnostic-related deaths, lower health-care costs and decrease the amount of time needed by radiologists to analyse the radiographs. Additionally, areas with minimal medical staff would greatly benefit from a model that effectively replaces radiologists.

References

- Data & Statistics | TB | CDC. (n.d.). Retrieved from <https://www.cdc.gov/tb/statistics/default.htm>
- Ittyachen, A. M., Vijayan, A., & Isac, M. (2017). The forgotten view: Chest X-ray - Lateral view. *Respiratory Medicine Case Reports*, 22, 257-259. doi:10.1016/j.rmcr.2017.09.009
- Johnson, A. E., Pollard, T. J., Berkowitz, S. J., Greenbaum, N. R., Lungren, M. P., Deng, C., . . . Horng, S. (2019). MIMIC-CXR: A Large Publicly Available Database of Labeled Chest Radiographs. Retrieved from <https://arxiv.org/abs/1901.07042>.
- "Lung Cancer Is the Biggest Cancer Killer in Both Men and Women" Infographic | CDC. (n.d.). Retrieved from https://www.cdc.gov/cancer/lung/basic_info/mortality-infographic.htm
- Pneumonia Can Be Prevented-Vaccines Can Help | CDC. (n.d.). Retrieved from <https://www.cdc.gov/pneumonia/prevention.html>
- Puddy, E., & Hill, C. (2007). Interpretation of the chest radiograph. *Continuing Education in Anaesthesia Critical Care & Pain*, 7(3), 71-75. doi:10.1093/bjaceaccp/mkm014
- Rajpurkar, P., Irvin, J., Ball, R. L., Zhu, K., Yang, B., Mehta, H., . . . Lungren, M. P. (2018). Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLOS Medicine*, 15(11). doi:10.1371/journal.pmed.1002686
- Raoof, S., Feigin, D., Sung, A., Raoof, S., Irugulpati, L., & Rosenow, E. C. (2012). Interpretation of Plain Chest Roentgenogram. *Chest*, 141(2), 545-558. doi:10.1378/chest.10-1302
- Rubin, J., Sanghavi, D., Zhao, C., Lee, K., Qadir, A., & Xu-Wilson, M. (2018). Large Scale Automated Reading of Frontal and Lateral Chest X-Rays using Dual Convolutional Neural Networks. Retrieved from <https://arxiv.org/abs/1804.07839>.

Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017). ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/cvpr.2017.369