Classification of chest radiographs using general purpose cloud-based automated machine learning: pilot study

Tamaghna Ghosh[1], Simran Tanwar[1], Shishir Chumber[2] and Kavita Vani[1*] (iD)

Abstract

Background: Widespread implementation of machine learning models in diagnostic imaging is restricted by dearth of expertise and resources. General purpose automated machine learning offers a possible solution.

This study aims to provide a proof of concept that a general purpose automated machine learning platform can be utilized to train a CNN to classify chest radiographs.

In a retrospective study, more than 2000 postero-anterior chest radiographs were assessed for quality, contrast, position, and pathology. A selected dataset of 637 radiographs were used to train a CNN using reinforcement learning based automated machine learning platform. Accuracy metrics of each label was calculated and model performance was compared to previous studies.

Results: The auPRC (area under precision-recall curve) was 0.616. The model achieved precision of 70.8% and recall of 60.7% ($P > 0.05$) for detection of "Normal" radiographs. Detection of "Pathology" by the model had a precision of 75.6% and recall of 75.6% ($P > 0.05$). The F1 scores were 0.65 and 0.75 respectively.

Conclusion: Automated machine learning platforms may provide viable alternatives to developing custom CNN models for classification of chest radiographs. However, the accuracy achieved is lower than a comparable traditionally developed neural network model.

Keywords: Computer-assisted diagnosis [E01.158], Radiography [E01.370.350.700], Neural networks [G17.485], Machine learning [G17.035.250.500], Cloud computing [L01.224.097], Deep learning [G17.035.250.500.250]

Background

Current scenario

Until recently, the approach to develop a CAD system to extract meaningful features and infer a diagnosis was based heavily on Rule Based algorithms [1]. These were hand crafted sets of definitions, which were used by the computing system to detect abnormalities. The levels of accuracy achieved by these systems were poor and remained an enhanced visualization function rather than an independent diagnostic tool [1].

Advances in Deep Learning algorithms have since surpassed the traditional Rule Based algorithms in accuracy [2]. Multiple Deep Learning algorithms have even been able to surpass human performance in sorting natural images [3, 4]. This has also led to interest in applying Deep Learning algorithms to diagnostic imaging and multiple studies have been done on application of Deep Learning on the interpretation of Chest radiographs. Litjens et al. produced an extensive survey of deep learning studies done on medical image datasets. It recorded 12 studies which had been undertaken till then, on application of deep learning techniques to chest radiographs to aid diagnosis [5].

Related works

The first published attempt at applying machine learning to this problem was by Lo et al. They designed a two-

* Correspondence: kavita2vani@gmail.com
[1]Department of Radiodiagnosis, Dr. Ram Manohar Lohia Hospital, Delhi, India
Full list of author information is available at the end of the article

layered convolutional neural network (CNN) to identify true pulmonary nodules on chest radiographs. The model was trained to differentiate between true nodules and end on vessels or rib overlap artifacts mimicking pulmonary nodules [6].

Anavi et al. created an image retrieval system that would rank the dataset images according to similarity with the query image. The model was created as a combination of pre-trained CNN with a support vector machine (SVM). The classification-based model was able to achieve a recall (recall of only top 30 images were reported) of 0.310 for left side pleural effusion, 0.182 for left side consolidation, and 0.103 for identification of a healthy chest [7, 8].

Multiple studies also focused on using CNN models which were previously trained with natural images, applying them to classification of chest x-rays in an attempt to reduce the complexity, cost, and time required. Bar et al. in a unique experiment used a pre-trained CNN and low level features to successfully detect lung pathologies. The Sensitivity of the model ranged from 0.80 to 0.89 and the specificity ranged from 0.79 to 0.87 [9, 10]. Cicero et al. trained and validated a GoogLe Net CNN model on a large data set of over 30,000 frontal chest radiographs to detect common lung pathologies. They were able to achieve a ROC AUC of 0.964 with a sensitivity and specificity of 91% for identification of Normal chest [11]. They also achieved a remarkable level of accuracy with AUC ranging from 0.850 to 0.962 for detection of common lung pathologies, with a sensitivity of 74 to 91% and specificity of 75 to 91% [11]. Hwang et al. designed a pre-trained fine-tuned 6 layered CNN which was able to process complete chest radiographs and was trained to detect pulmonary tuberculosis. The average ROC AUC of the model without transfer learning was 0.816 and with transfer learning it improved to 0.964 [12]. Wang et al. used a ImageNet pre-trained CNN along with hand crafted feature selection to identify pulmonary nodules on chest radiographs. They were able to achieve a sensitivity of 69.27% and specificity of 97.02% [13].

Some studies have attempted to overcome the exponential rise in algorithm complexity that accompanies the quest for greater accuracy by deploying multiple synergistically working trained machine learning models. Shin et al. produced a highly sophisticated model with a CNN predicting the lung pathology and a Recurrent Neural Tensor Network providing short captions for annotation. The model was trained on a large dataset of 7000 images [14]. Islam et al. trained and tested multiple deep learning models to detect lung pathologies on frontal chest radiographs. Their experiments proved that different models excelled at single specific pathology detection and using ensembles of models improved the overall performance [1]. Wang et al. created the first standardized public dataset of chest radiographs to develop a benchmark against which all deep network performances could be assessed [15]. It also tested several different standard CNN models against the dataset. The average ROC AUC for the best performing model was 0.738. Yao et al. developed a novel method to leverage associations of different lung pathologies to improve detection accuracy in their model. The model outperformed Wang et al. with an average ROC AUC of 0.798 [16]. Rajpurkar et al. developed CheXNet, a 121 layered CNN which was trained on Chest X ray -14 public dataset for detection of signs of pneumonia on chest radiographs [17]. The model outperformed the published results of both Wang et al. and Yao et al. on all listed pathologies of the dataset. Lakhani and Sundaram trained an ensemble of GoogLeNet and AlexNet DCNNs to detect pulmonary tuberculosis. They achieved an impressive ROC AUC of 0.99 with a sensitivity of 97.3% and specificity of 100% [18].

There has been simultaneous development in Deep Neural Network models that aid in other challenging aspects of chest radiograph interpretation. Kim and Hwang developed an ML framework to detect tuberculosis by projecting heat maps on the suspicious areas of chest radiographs [19]. Rajkomar et al. created a pre-trained DCNN model to sort chest radiographs into anteroposterior and lateral views [20]. Yang et al. used a cascading set of CNNs to detect and suppress bone from standard chest radiographs to render clear view of the pulmonary and cardiac soft tissue shadow [21].

Problem statement

While the potential of neural networks to change the diagnostic imaging is clear, the real world application of such research remains greatly hindered due to the prohibitive cost of running multiple graphics processing units (GPUs) to train such neural networks. While the cost of such technology is reducing, it still remains in the range of $100,000 [2]. An additional obstacle for accessing and applying machine learning techniques in radiology is the lack of expertise and knowledge required for hyperparameter tuning, data augmentation, etc. It also remains a time-consuming process as the complexity of the machine learning model increases combinatorially and requires considerable experimentation even by those with machine learning expertise. Recent advances have attempted to automate the designing process of machine learning models by evolutionary algorithms and reinforcement learning algorithms [22, 23]. Multiple proprietary Application Programming Interfaces (APIs) for automated machine learning based on reinforcement learning are now available. These offer the ability to access and train a neural network for a fractional cost and

Fig. 1 Flowchart summarizing the data categorization and model development process
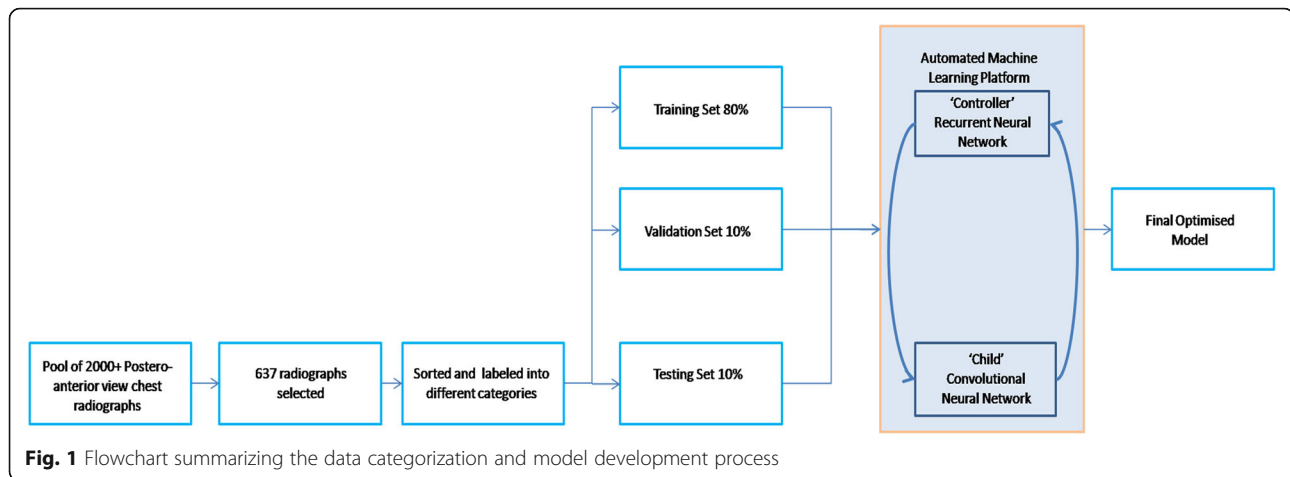
time to that of traditional machine learning models. This is critical for democratizing the access to machine learning and universalizing the use of this technology. It is especially important considering the potential beneficial outcomes of such research—reducing cost of diagnostic imaging, streamlining work flows, and enabling greater penetration of diagnostic imaging to the community level—are most required in resource limited developing communities.

This pilot study was aimed to provide a proof of concept that general purpose automated machine learning platforms such as Google AutoML Vision can be utilized to train a neural network to diagnose and categorize chest radiographs in a real-world setting.

Methods

Dataset creation

A pool of over 2000 postero-anterior view chest radiographs from the out-patient and in-patient department acquired on different computed radiography and digital radiography systems were assessed for quality, level of penetration, positioning, and contrast. Those with very poor quality, low contrast, and unsatisfactory positioning were rejected. However, chest radiographs with minor imperfections in breath-holding, positioning or contrast, deemed reportable by the radiologist were included. Chest radiographs with clothing, jewelry, and implantable medical devices artifacts were included to mirror real-world variations.

Fig. 2 Distribution of conspicuousness of pathology in the training data set
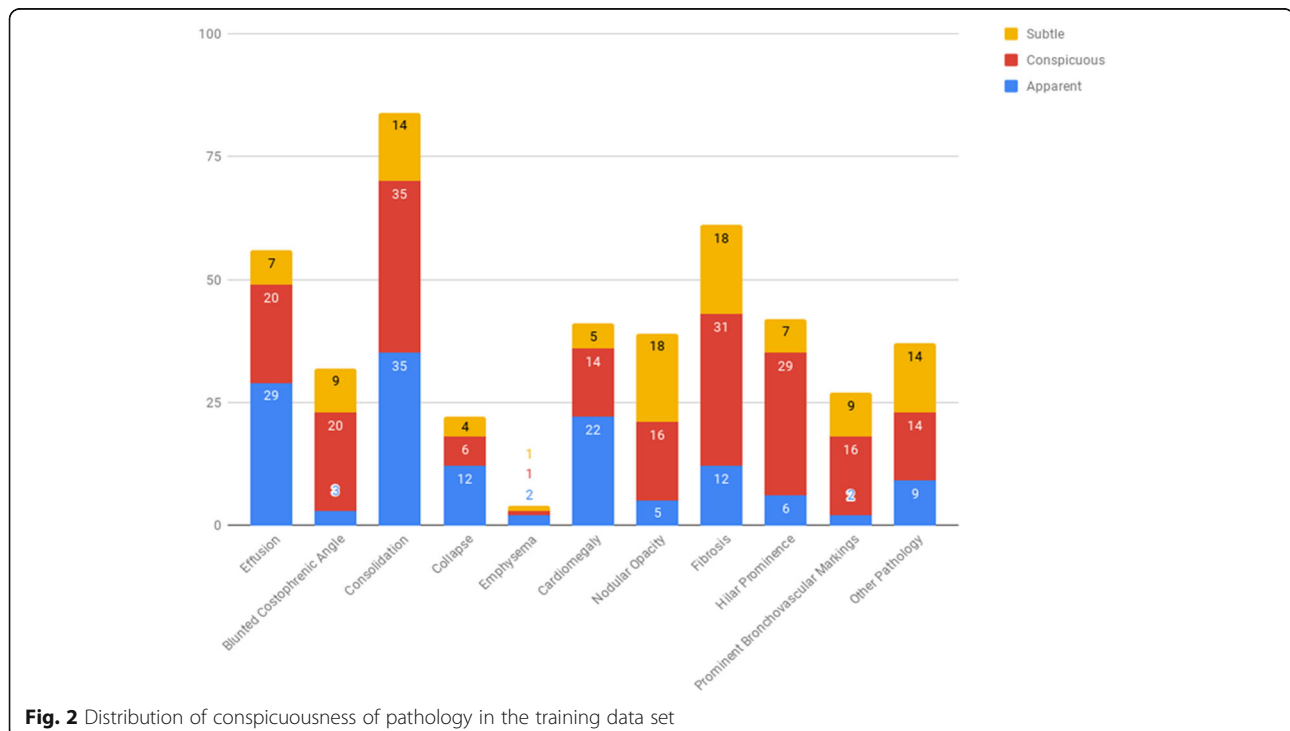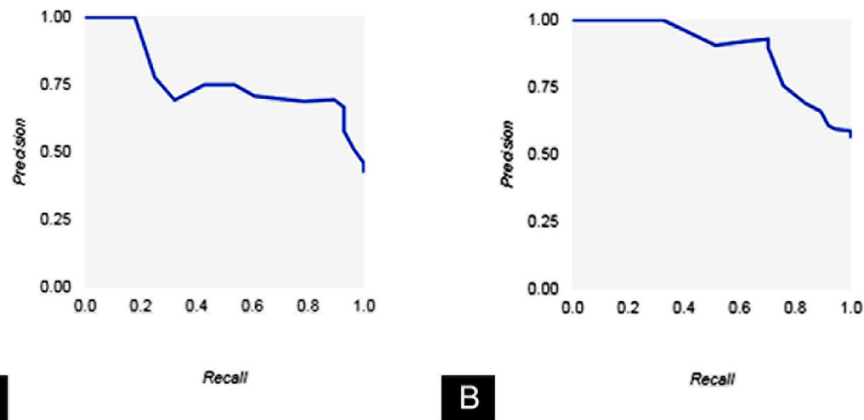
Fig. 3 Precision-recall tradeoff graphs for classification of "Normal" (**a**) and "Pathology" (**b**)

Image processing

The resultant dataset of 637 images were then converted from proprietary file types into Joint Photographic Expert Group file type with a 1024 × 1024 matrix size with 96 dpi vertical and horizontal resolution encoded using baseline DCT Huffman coding. The bit depth (bits per sampling) was set at 8bits with a Chroma subsampling Y'CbCr=4:2:0. The dataset was de-identified and was compliant with the Health Insurance Portability and Accountability Act. No data augmentation procedures were performed.

Model implementation

The dataset was uploaded onto Google Cloud Platform (Google LLC, Menlo Park, CA, USA) and processed using Cloud AutoML Vision Beta (release date: July 24,

2018). Multiple labels were created for classifying different pathologies and image characteristics (Fig. 1). Each image was labeled with one or more labels using Vision UI running on Chrome v68.0.3440 (Google LLC, Menlo Park, CA, USA). The dataset was randomly subdivided with 80% of the images under each category allocated to the training set and 10% each to the validation and testing sets. The system used had Intel Core i3 – 4005U 1.70 GHz chipset (Intel, Santa Clara, CA, USA), 4.00 Gb RAM, 512 Gb hard disk space, integrated Intel HD Graphics 4400 (Intel, Santa Clara, CA, USA) with Windows 7 Ultimate Operating System (Microsoft Corporation, Redmond, WA, USA).

Statistical analysis was performed with the in-built metrics projection in Vision API, Google Sheets (Google LLC, Menlo Park, CA, USA) and MedCalc (MedCalc Software Ltd, Ostend, Belgium).

Results

Dataset characteristics

The dataset contained 637 postero-anterior view chest radiographs of which 332 had some pathology (52.1%). The dataset had a mild male predominance (57.8%) with an average age of 26.5 years. Each image assessed subjectively for quality and marked either satisfactory or poor. 82.1% of the images were of satisfactory quality but the dataset also contained 17.9% radiographs which were poor in quality but still deemed reportable by the radiologist. 47.6% of the dataset contained some form of artifact from clothing, jewelry, or implantable devices like pacemakers. The images were also assessed for positioning of the subject and revealed 25.9% to have some degree of rotation—which could lead to certain artifactual findings such as apparent cardiomegaly and prominence of the hila. Forty-three of the 637 radiographs were found to have been acquired in mid-inspiration. These imperfect images were introduced

Table 1 Accuracy metrics

Label	Number of images	Precision	Recall
Normal	305	70.8%	60.7%
Pathology	332	75.6%	75.6%
Artifact	303	55.3%	63.6%
Cardiomegaly	43	100%	0%
Collapse	21	100%	0%
Consolidation	84	100%	0%
Costophrenic_Angle_Blunted	31	100%	0%
Fibrosis	61	100%	0%
Hilar_Prominence	42	100%	0%
Midinspiratory	43	100%	0%
Nodular_Opacities	39	100%	0%
Other_Pathology	37	100%	0%
Pleural_Effusion	56	100%	0%
Prominent_BV_Markings	27	100%	0%
Rotated	165	33.3%	5.9%

Table 2 Evaluation of model performance in detecting "Normal" chest radiographs

	Normal	Not normal	Total
Classified as "Normal"	185 (true positive)	76 (false positive)	261
Not classified as "Normal"	120 (false negative)	256 (true negative)	376
Total	305	332	
Statistics	**Value**	**95% confidence interval**	
Sensitivity	60.66%	54.93 to 66.17%	
Specificity	77.11%	72.21 to 81.52%	
Positive likelihood ratio	2.65	2.13 to 3.29	
Negative likelihood ratio	0.51	0.44 to 0.59	
Positive predictive value	70.88%	66.21 to 75.15%	
Negative predictive value	68.09%	64.71 to 71.28%	
Accuracy	69.23%	65.48 to 72.80%	

into the dataset to reduce overfitting of the model to the training set and improve its real world applicability.

The images with pathology were sub-classified and labeled into 9 different categories (Fig. 2). The pathologies were also assessed for subjective conspicuity. Each lung field was divided into three lung zone: upper, middle, and lower. A pathology occupying more than or equal to half of a zone was deemed "Apparent." If the pathology occupied less than half but more than 25% of the lung zone, it was marked as "Conspicuous." Lesions occupying less than 25% of a lung zone were termed "Subtle." The distributions of the lesions are shown in Fig. 2.

Accuracy metrics

The precision (positive predictive value) for all labels was 65.7% with a recall (sensitivity) of 40.1%. The auPRC (area under precision-recall curve or average precision) of the model was 0.616 (Fig. 3). The precision and recall for each category is summarized in Table 1. The F1 Score for classification was 0.65 for "Normal" category and 0.75 for "Pathology" category. Further evaluation statistics for both categories are summarized in Tables 2 and 3, respectively.

Discussion

Unmet needs

While there has been considerable interest in the application of convolutional neural networks and other forms of machine learning for classification of chest radiographs into various pathologies, the underlying technology utilized in all these studies remain exclusionary [5, 24–26]. These studies either constructed and trained machine learning models de novo or worked with pre-trained CNNs like AlexNet and GoogLeNet [3, 27]. Though these methods yielded high accuracy models which could classify chest pathologies, they were built on systems which required high level of expertise as well as prohibitively costly infrastructure. This has led to a data-algorithm divide. The predictive accuracy of an algorithm is strictly contingent on the dataset that it is trained on (Fig. 4). But a large number of institutions in resource-limited settings may not have access to

Table 3 Evaluation of model performance in detecting "Pathology" chest radiographs

	Pathology	Not pathology	Total
Classified as "Pathology"	251 (true positive)	81 (false positive)	332
Not classified as "Pathology"	81 (false negative)	224 (true negative)	305
Total	332	305	
Statistics	**Value**	**95% confidence interval**	
Sensitivity	75.60%	70.62 to 80.13%	
Specificity	73.44%	68.11 to 78.31%	
Positive likelihood ratio	2.85	2.34 to 3.46	
Negative likelihood ratio	0.33	0.27 to 0.41	
Positive predictive value	75.60%	71.80 to 79.04%	
Negative predictive value	73.44%	69.34 to 77.18%	
Accuracy	74.57%	71.00 to 77.91%	
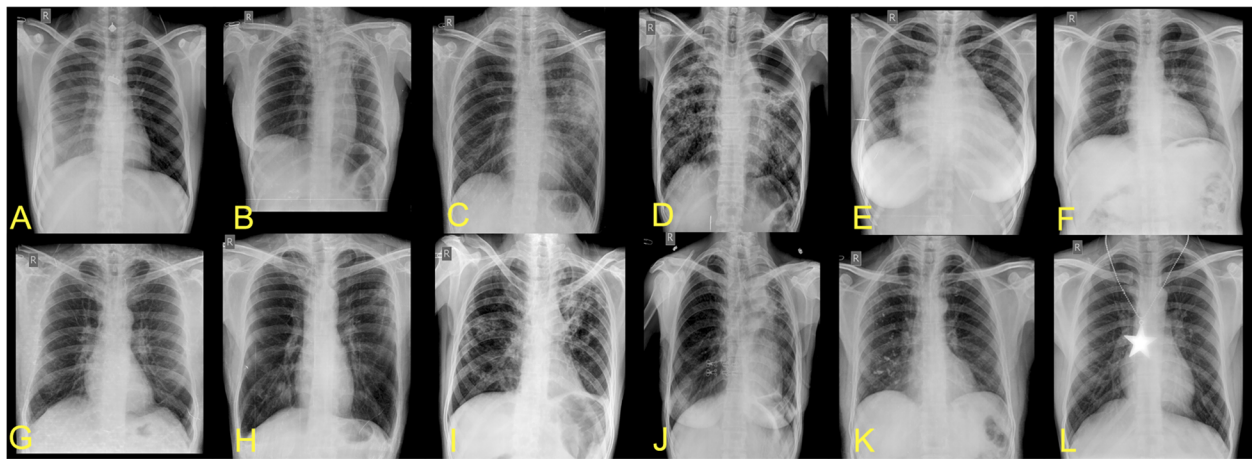
Fig. 4 Examples from the training data set. Right-sided pleural effusion (**a**). Fibrosis in left upper zone (**b**). Consolidation in bilateral lung fields (**c**). Multiple pathologies in single radiograph—showing patchy consolidation with fibrotic bands and bulla in left upper zone (**d**). Cardiomegaly (**e**). Left hilar prominence (**f**). Prominent bronchovascular markings (**g**). Emphysema (**h**). Blunted left costophrenic angle with patchy consolidation (**i**). Collapse of left upper and middle zone (**j**). Nodular opacities in right lower zone (**k**). Normal chest radiograph with clothing artifact (**l**)

machine learning technology but do have access to large volumes of data.

Proposed solution

In this study, we tried to explore the possibility of repurposing general purpose automated machine learning models to classify diagnostic images, in particular chest radiographs. The platform used was Cloud AutoML Vision, which circumvents the challenges of requiring a large amount of time and expertise in crafting a neural network by using reinforcement learning [23]. The "controller" recurrent network creates variable length strings. These strings act as templates for development of "child" convolutional neural networks. These "child" networks are trained on the dataset and subsequently evaluated for accuracy. The accuracy metric is used as a positive reinforcement for the "controller" network. Thus in the subsequent iteration, the "child" networks with higher accuracy are favored. This is repeated until a single best "child" network is achieved with the highest accuracy.

Model accuracy

The accuracy metric of our trained model was expectedly lower than dedicated CNNs. The model had very poor sensitivity for sub-classification of pathology. However, the overall accuracy achieved for detection of pathology in chest radiographs was 74.57%. The accuracy parameters of the model are compared with two studies conducted with comparable machine learning models in Table 4.

Our model, DeepDx, was able to achieve comparable accuracy to the model used by Bar et al., even surpassing their precision rate by almost 25%. This is substantial progress, especially when viewed in the context of the highly specialized fusion model (two separate deep learning baseline descriptors used along with GIST descriptor) created by Bar et al. [10]. The comparison table also reveals that Cicero et al. in their study achieved a much higher overall accuracy, but the success could be attributed at least in part to the large dataset on which their model was trained [11].

Justification

In our model, the three categories with examples above the minimum recommended number did provide good accuracy and with targeted increase in the dataset in subsequent iterations the overall model accuracy is likely to improve further. As per documentation released with Cloud AutoML Vision (Google LLC, Menlo Park, CA, USA), which we utilized in the study, the minimum recommended examples per label is 100 and approximately 1000 examples are advised for accurate prediction. This may not always be feasible for medical imaging, as rarity is often a feature of diseases with serious implications; and the time required to accrue enough examples may impede progress. This problem is usually circumvented by data augmentation procedures. Application of techniques such as horizontal flipping, cropping, rotation, and padding on chest radiographs and their effect on

Table 4 Comparison of accuracy metrics

	Dataset	Precision	Recall	Specificity
DeepDx	637	75.60%	75.60%	73.44%
Bar et al.	443	51.25%	84.00%	78.00%
Cicero et al.	38101	90.00%	91.00%	91.00%
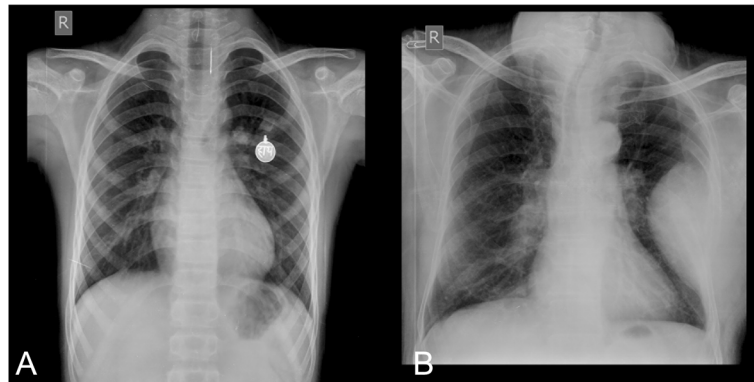
Fig. 5 a False positive—the algorithm misinterpreted the artifact as nodular opacity. **b** False negative—the algorithm failed to identify the pleural based mass as "Pathology"

model accuracy has not been investigated. It may not be prudent to shoehorn techniques, while efficacious in other image datasets, onto diagnostic images. For example, horizontal flipping of a chest radiograph may create false-positive results for detection of cardiomegaly and may in fact reduce accuracy. The model may also fail to flag cases of dextrocardia. Similarly, training machine learning models on rotated radiographs may lead to the algorithms assigning undue importance to irrelevant components of the image. Also many disease processes are defined by their orientation like cephalization of vessels in CCF, which may be lost during the augmentation process.

Accuracy of the model is also likely to gain from changing the labeling structure of the dataset. In our study, we trained the algorithm to diagnose "Normal" and "Pathology" not as a binary alternative but as distinct classification categories. This was done keeping in mind real world application, as many radiographs do not distinctly fit either into an apparently normal or disease category. Many radiographs have suspicious features which should not be classified as disease and may require consensus reporting by radiologists. Another advantage of detecting the two categories separately was that it gave us comparable statistics with a larger number of studies, as most have trained to classify either one of the categories. The downside of this labeling structure was that it added to the complexity and thus probably reduced accuracy of the model. The model can be trained, in further studies; to detect only "Pathology" and the "Normal" can be processed as a default class. The sensitivity of the "Pathology" label should be increased to commit false positives and catch the indeterminate cases rather than being labeled "Normal" (Fig. 5). This will again entail human intervention to sort through and weed out the false positive, but will improve accuracy.

Reflections

The study has highlighted certain definite advantages of using automated machine learning in developing diagnostic classification models. The method reduces infrastructure requirements and cost to a fractional amount. The ease of use, with GUIs, also enables implementation and fine tuning without cumbersome coding languages. The reinforcement-based learning model greatly reduces the time requirement for developing complex CNN architecture. And importantly, such platforms provide scalability to improve upon a model and add further complexity to the classifier.

Future implications

Further work needs to be done with larger datasets of diagnostic images, to ascertain the maximal overall accuracy achievable. Multiple platforms now exist providing similar tools and they should be evaluated in a controlled trial for unbiased comparison. Data augmentation procedures should also be validated for use with medical imaging, particularly radiological images. Lastly, most studies attempting to classify chest radiographs have dealt with post-processed compressed images converted to non-native file types such as JPEG and PNG [11, 12, 17, 18]. This conversion may lead to loss of important image characteristics and attempts should be made to use DICOM file types for future training of algorithms.

Conclusion

Computer vision is revolutionizing the field of diagnostic imaging. But its resource-intensive nature may preclude its wider implementation and acceptance. This study presented an alternative to the traditional machine learning infrastructure and aimed to investigate the use of commercially available general purpose cloud-based automated machine learning for detection of pathologies on standard postero-anterior chest radiographs.

The study found automated machine learning to be a viable alternative to human designed diagnostic convolutional neural networks. The accuracy of the model developed was conservative in comparison to standard deep learning models. However, restructuring of the classifiers and increasing the training dataset hold promise of achieving greater accuracy. Further multi-platform studies are required with larger datasets to fully explore its potential.

While machine learning promises of vast improvements in speed and accuracy in detection of pathologies across imaging modalities, greater research focus needs to be directed towards ensuring that this novel technology is used to bridge the health-wealth gap and not widen it.

Abbreviations
CNN: Convolutional neural networks; auPRC: Area under precision-recall curve; CAD: Computer-aided diagnosis; AUC: Area under the curve; ROC: Receiver-operator curve; SVM: Support vector machine; ML: Machine learning; GPU: Graphics processing nit; API: Application Programming Interface; GIST: General Information on Shape and Texture; GUI: Graphic user interface; JPEG: Joint Photographic Expert Group; PNG: Portable network graphics

Authors' contributions
TG, ST, and SC collected and analyzed the clinical data. TG, ST, and SC wrote the manuscript. KV guided, supervised, and edited the manuscript. All authors read and approved the final manuscript.

Availability of data and materials
YES, The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate
The need for approval was waived as retrospective analysis of blinded data ensuring anonymity of patients was done for study.

Consent for publication
Retrospective evaluation of blinded digital data was done ensuring anonymity of patients.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Radiodiagnosis, Dr. Ram Manohar Lohia Hospital, Delhi, India. [2]Department of Neurology, Dr. Ram Manohar Lohia Hospital, Delhi, India.

References
1. Islam MT, Aowal MA, Minhaz AT, Ashraf K (2017) Abnormality detection and localization in chest x-rays using deep convolutional neural networks. arXiv arXiv:1705.09850
2. Dreyer KJ, Geis JR (2017) When machines think: radiology's next frontier. Radiology 285(3):713–718. https://doi.org/10.1148/radiol.2017171183
3. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp 1097–1105
4. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
5. Litjens G, Kooi T, Bejnordi BE, Setio AA, Ciompi F, Ghafoorian M et al (2017) A survey on deep learning in medical image analysis. Med Image Anal 42: 60–88. https://doi.org/10.1016/j.media.2017.07.005
6. Lo SC, Lou SL, Lin JS, Freedman MT, Chien MV, Mun SK (1995) Artificial convolution neural network techniques and applications for lung nodule detection. IEEE Trans Med Imaging 14(4):711–718. https://doi.org/10.11 09/42.476112
7. Anavi Y, Kogan I, Gelbart E, Geva O, Greenspan H (2015) A comparative study for chest radiograph image retrieval using binary texture and deep learning classification. In: Engineering in medicine and biology society (EMBC), 2015 37th annual international conference of the IEEE. IEEE, New York, pp 2940–2943
8. Anavi Y, Kogan I, Gelbart E, Geva O, Greenspan H (2016) Visualizing and enhancing a deep learning framework using patients age and gender for chest x-ray image retrieval. In: Tourassi GD, Armato SG III (eds) Medical imaging 2016: computer-aided diagnosis. SPIE, New York, p 978510. https://doi.org/10.1117/12.2217587
9. Bar Y, Diamant I, Wolf L, Lieberman S, Konen E, Greenspan H (2015) Chest pathology detection using deep learning with non-medical training. In: 2015 IEEE 12th international symposium on biomedical imaging (ISBI). IEEE, New York, pp 294–297. https://doi.org/10.1109/ISBI.2015.7163871
10. Bar Y, Diamant I, Wolf L, Greenspan H (2015) Deep learning with non-medical training used for chest pathology identification. In: Medical imaging 2015: computer-aided diagnosis. SPIE, Orlando, p 94140. https://doi.org/10.1117/12.2083124
11. Cicero M, Bilbily A, Colak E, Dowdell T, Gray B, Perampaladas K, Barfett J (2017) Training and validating a deep convolutional neural network for computer-aided detection and classification of abnormalities on frontal chest radiographs. Investig Radiol 52(5):281–287. https://doi.org/10.1097/RLI.0000000000000341
12. Hwang S, Kim HE, Jeong J, Kim HJ (2016) A novel approach for tuberculosis screening based on deep convolutional neural networks. In: Tourassi GD, Armato SG III (eds) Medical imaging 2016: computer-aided diagnosis. SPIE, New York, p 97852. https://doi.org/10.1117/12.2216198
13. Wang C, Elazab A, Wu J, Hu Q (2017) Lung nodule classification using deep feature fusion in chest radiography. Comput Med Imaging Graph 57:10–18. https://doi.org/10.1016/j.compmedimag.2016.11.004
14. Shin HC, Roberts K, Lu L, Demner-Fushman D, Yao J, Summers RM (2016) Learning to read chest x-rays: recurrent neural cascade model for automated image annotation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, Las Vegas, pp 2497–2506. https://doi.org/10.1109/CVPR.2016.274
15. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM (2016) 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. IEEE, Honolulu, pp 3462–3471. https://doi.org/10.1109/CVPR.2017.369
16. Yao L, Poblenz E, Dagunts D, Covington B, Bernard D, Lyman K (2017) Learning to diagnose from scratch by exploiting dependencies among labels. arXiv arXiv:1710.10501
17. Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T et al (2017) Chexnet: radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv arXiv:1711.05225
18. Lakhani P, Sundaram B (2017) Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology 284(2):574–582. https://doi.org/10.1148/radiol.2017162326
19. Kim HE, Hwang S (2016) Deconvolutional feature stacking for weakly-supervised semantic segmentation. arXiv arXiv:1602.04984
20. Rajkomar A, Lingam S, Taylor AG, Blum M, Mongan J (2017) High-throughput classification of radiographs using deep convolutional neural networks. J Digit Imaging 30(1):95–101. https://doi.org/10.1007/s10278-016-9914-9

21. Yang W, Chen Y, Liu Y, Zhong L, Qin G, Lu Z et al (2017) Cascade of multi-scale convolutional neural networks for bone suppression of chest radiographs in gradient domain. Med Image Anal 35:421–433. https://doi.org/10.1016/j.media.2016.08.004

22. Real E, Moore S, Selle A, Saxena S, Suematsu YL, Tan J et al (2017) Large-scale evolution of image classifiers. In: Proceedings of the 34th international conference on machine learning-volume 70, pp 2902–2911

23. Bello I, Zoph B, Vasudevan V, Le QV (2017) Neural optimizer search with reinforcement learning. In: Proceedings of the 34th international conference on machine learning-volume 70, pp 459–468

24. Sim Y, Chung MJ, Kotter E, Yune S, Kim M, Do S, Han K, Kim H, Yang S, Lee DJ, Choi BW (2020) Deep convolutional neural network–based software improves radiologist detection of malignant lung nodules on chest radiographs. Radiology 294(1):199–209. https://doi.org/10.1148/radiol.2019182465

25. Park S, Lee SM, Kim N, Choe J, Cho Y, Do KH et al (2019) Application of deep learning–based computer-aided detection system: detecting pneumothorax on chest radiograph after biopsy. Eur Radiol 29(10):5341–5348. https://doi.org/10.1007/s00330-019-06130-x

26. Dunnmon JA, Yi D, Langlotz CP, Ré C, Rubin DL, Lungren MP (2019) Assessment of convolutional neural networks for automated classification of chest radiographs. Radiology 290(2):537–544. https://doi.org/10.1148/radiol.2018181422

27. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D et al (2015) Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, Boston, pp 1–9

Publisher's Note