Classification of COVID-19 in chest radiographs: assessing the impact of imaging parameters using clinical and simulated images

Fricks, Rafael, Abadi, Ehsan, Ria, Francesco, Samei, Ehsan

Rafael Fricks, Ehsan Abadi, Francesco Ria, Ehsan Samei, "Classification of COVID-19 in chest radiographs: assessing the impact of imaging parameters using clinical and simulated images," Proc. SPIE 11597, Medical Imaging 2021: Computer-Aided Diagnosis, 115970A (15 February 2021); doi: 10.1117/12.2582223

SPIE.

Event: SPIE Medical Imaging, 2021, Online Only

Classification of COVID-19 in Chest Radiographs: Assessing the Impact of Imaging Parameters using Clinical and Simulated Images

Rafael B. Fricks[a,b], Ehsan Abadi[b], Francesco Ria[b], Ehsan Samei[b]
[a]National Artificial Intelligence Institute, U.S. Department of Veterans Affairs, 810 Vermont Ave NW, Washington D.C., USA 20420;
[b]Carl E. Ravin Advanced Imaging Laboratories, Duke University, 2424 Erwin Rd, Durham, NC, USA 27705

ABSTRACT

As computer-aided diagnostics develop to address new challenges in medical imaging, including emerging diseases such as COVID-19, the initial development is hampered by availability of imaging data. Deep learning algorithms are particularly notorious for performance that tends to improve proportionally to the amount of available data. Simulated images, as available through advanced virtual trials, may present an alternative in data-constrained applications. We begin with our previously trained COVID-19 x-ray classification model (denoted as CVX) that leveraged additional training with existing pre-pandemic chest radiographs to improve classification performance in a set of COVID-19 chest radiographs. The CVX model achieves demonstrably better performance on clinical images compared to an equivalent model that applies standard transfer learning from ImageNet weights. The higher performing CVX model is then shown to generalize effectively to a set of simulated COVID-19 images, both quantitative comparisons of AUCs from clinical to simulated image sets, but also in a qualitative sense where saliency map patterns are consistent when compared between sets. We then stratify the classification results in simulated images to examine dependencies in imaging parameters when patient features are constant. Simulated images show promise in optimizing imaging parameters for accurate classification in data-constrained applications.

Keywords: Classification, Deep Learning, DenseNet, COVID-19, Radiography, Quantitative Imaging, Virtual Clinical Trials, XCAT Phantom

1. INTRODUCTION

When new diseases such as COVID-19 emerge, data collected in the course of treatment generally contributes to improvements in the standard of care. This process is underway in the current pandemic, where early observations led to improved guidelines for diagnosis and disease management via imaging.[1] As the pandemic unfolds, data collection has escalated with national mobilization to curate archives of imaging data on COVID-19, such as the National COVID Cohort Collaborative[2] or the Medical Imaging and Data Resource Center.[3] Ultimately large, high-quality datasets allow researchers to bring notoriously data-hungry artificial intelligence tools to bear on improving patient outcomes, as happened in chest radiography prior to COVID-19,[4,5] or lung cancer screening[6–8] and management[7] with CT. While such massive databases are the gold standard data source for machine learning research, their collection and annotation requires a concerted effort often spanning years or decades. This timeline is too costly for COVID-19.

We investigate a supplement to enhance the traditional course of data collection and accelerate applications of machine learning to new disease. From decades of development, advances in Virtual Imaging Trial (VIT) tools have yielded realistic phantoms combined with the ability to accurately simulate imaging systems. These tools allow researchers to specify a pathology based on currently understood imaging characteristics of the disease. In the resulting images, the disease state is known exactly, and the same simulated patient can be imaged under various imaging protocols. Reasonably realistic simulated images could therefore be used to examine the effect of imaging parameters on algorithmic classification, with a constant underlying patient and pathological state.

Using VIT tools, we generated a series of 120 simulated radiographs from 30 XCAT phantoms, where two-thirds of patients expressing COVID-19 signatures. We applied a convolutional neural network, previously trained[9] on existing chest x-ray databases[4,10] as well as a new set of 1060 chest x-rays of RT-PCR confirmed clinical COVID-19 cases, to classify the simulated images. Evaluating the results via the analysis of the receiver operating characteristic curves (ROC)

Medical Imaging 2021: Computer-Aided Diagnosis, edited by Maciej A. Mazurowski, Karen Drukker, Proc. of SPIE Vol. 11597, 115970A · © 2021 SPIE
CCC code: 1605-7422/21/$21 · doi: 10.1117/12.2582223

for clinical and simulated images, we found comparable quantitative performance between the clinical and simulated images. Qualitative results produced via Grad-CAM[11] visualizations of network activation showed similar detection pattern in both populations. Ongoing work includes stratifying the results based on x-ray dose and patient characteristics, to examine how to individualize imaging protocols for maximum classification accuracy. Results imply simulated data may be used to augment development of machine learning algorithms for more proficient and timely classification of COVID-19.

2. METHODS

2.1 Deep convolutional neural network for classifying COVID-19

In a previous study,[9] we trained a convolutional neural network for COVID-19 classification using 1060 radiographs collected from 475 patients, where 801 images come from patients with COVID-19 diagnoses confirmed by reverse-transcriptase polymerase chain reaction (RT-PCR) laboratory tests. The source images were manually cropped to remove major burn-in labels and unnecessary field of view, particularly in mobile x-ray acquired from patients in a supine position which were present in this dataset.

Our classification network employed the DenseNet121[12] architecture, initialized with ImageNet[13] weights, and was trained in three phases with for gradual domain adaptation[14]. In preprocessing, images were rescaled to a 512x512 resolution and 8-bit pixel depth and normalized with mean and standard deviation values estimated from a subset of images. At runtime, images were presented in batches of 16, augmented with random horizontal flipping, random rotation of up to 8 degrees, and random translation and cropping of up to 10% of the image size. Training employed a stochastic gradient descent with momentum optimizer, with initial learning rate of 0.01, momentum of 0.9, and weight decay of 0.0001. The network was trained to minimize the unweighted binary cross-entropy loss, with learning rate reduction by a factor of 10 at loss plateaus, and early stopping when performance did not improve after ten epochs. At each phase, the optimizer was reinitialized with a lower initial learning rate (0.01, 0.001, 0.0005).

In each training phase, the data source is modified to gradually adapt the network for COVID-19 classification. The network is first trained for a standard multilabel task[5] in chest radiography using the ChestX-ray14 database[4], which contains a variety of common chest x-ray findings. In the second phase, the network is further refined for pneumonia classification on a pneumonia Kaggle challenge dataset[10], which has higher pneumonia prevalence compared to the ChestX-ray14 database (2.18% vs 24.05%). Finally, the network is trained and evaluated using the collected COVID-19 clinical images. The images are split at the patient level into a training set of 626 images, a validation set of 216 images, and a test set of 218 images. Sampling is stratified to maintain roughly 75% COVID-19 prevalence as in the original set.

To quantify the performance of our COVID-19 X-ray classifier on clinical data (denoted as CVX), we compare the performance benefit of the additional training phases to a more standard transfer learning approach that omits the first two phases (denoted as STL). We plot the receiver operating characteristic curves for both classifiers on the COVID-19 test set and compute the area under the ROC curve (AUC). Additionally, we compare the model AUCs using the fast implementation[15] of DeLong's method,[16] which is used to test for significance as well as computing the confidence intervals for the estimated AUC. The model predictions are further explored qualitatively, using Grad-CAM[11] to generate saliency maps for each image that highlight image areas based on their contribution to the overall classification decision.

2.2 Simulating radiographs of COVID-19 cases

For simulating images, we used the 4D extended cardiac-torso (XCAT) phantoms, highly detailed human models developed in our laboratory. Previous extensions to XCAT have expanded its ability to represent various patient demographics and attributes, including sex, age, height, and weight. These variations were based on observed clinical statistics and patient data. Thousands of anatomical structures were represented with statistically modeled tissue heterogeneity and include modeled cardiac and respiratory motions. Known COVID-19 imaging findings, such as ground-glass opacities or consolidations were added to represent pathological states.[17]

For simulating radiography, we performed x-ray simulation in DukeSim.[18] DukeSim used ray-tracing computations to model the primary signal in radiography systems, with specifications matching a Siemens radiography system. Additionally, scatter estimations were generated through Monte Carlo simulations.[19]

In this study, 30 XCAT phantoms (20 COVID-19 and 10 normal incorporated in 10 XCAT models) were used. Each phantom was imaged using DukeSim at 4 different current-time product levels ranging from 0.5 to 4.0 mAs, stepping by

factors of 2. The patients range in diameter from 21.2 cm to 30.0 cm, representing various habitus. All simulated images were preprocessed similarly to clinical images, including rescaling of the image resolution to 512x512 and pixel range to an 8-bit representation, with normalization based on batch-estimated values.

Given this set of 120 simulated images under various imaging conditions, the two previous models are used to make predictions on the simulated data set. We plot ROC curves for model predictions made on the set of simulated COVID-19 chest radiographs and compute the AUC. Estimated AUCs are compared using DeLong's method.[16] We also generate saliency maps for each simulated image to qualitatively evaluate which features are being incorporated into the model decision. We examine classification performance on select cases where typical COVID-19 imaging findings are added to otherwise healthy images, to compare which regions the classifier weighs in its decision and whether those correspond to lesion locations.

2.3 Evaluation of Imaging Conditions

As a preliminary investigation into the impact of imaging parameters on individual characteristics, we plotted the ROC for each exposure level measured in mAs. Further driving into the impact on classification performance, we selected an operating threshold that approximately maximizes the F1 score of the full simulated image set ROC and examine the misclassifications as a function of patient diameter and radiation dose.

3 RESULTS

3.1 Classifier performance on clinical COVID-19 images

Evaluating the performance of the COVID-19 X-ray classifier (CVX) compared to a more standard transfer learning (STL) approach, there is a noticeable improvement in classification performance on clinical images of COVID-19 patients. Figure 1 plots receiver operating characteristic curves (ROC) for both models. The CVX network is pretrained with a specific curriculum to successively refine the classifier for the desired task. CVX is refined first on general chest radiographic findings, then refined to classify a set with higher pneumonia prevalence, then finally refined for COVID-19 detection. The resulting CVX model achieves an AUC of 0.849 (95% CI: 0.786 – 0.912), compared to the AUC for a more standard transfer learning approach of 0.757 (95% CI: 0.683 – 0.831). At $\alpha = 0.05$, there is a statistically significant difference between the performance of these two models as evaluated by DeLong's method for non-parametric comparison of two classifiers, with a calculated p-value of 0.047.

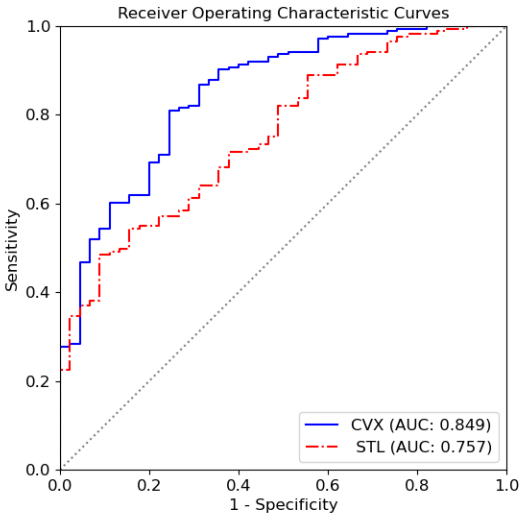

Figure 1. Comparison of receiver operating characteristic curves (ROC) for the COVID-19 X-ray classifier (CVX) to a standard transfer learning approach (STL). The CVX classifier is sequentially pretrained with data on similar tasks, presented to the network in a deliberate curriculum to enhance COVID-19 classification performance. While the standard transfer learning approach performs adequately on this data set, there is a statistically significant improvement in CVX's performance at $\alpha = 0.05$.

Examining the Grad-CAM activation maps for the CVX network, there is a consistent weighting pattern associated with high activation values at the network output (Figure 2). A jet colormap is used to represent network activations in their approximate spatial coordinates, with high-weight areas in red and low-weight areas in blue. Focal opacities in the lung (red) which appear to be consolidations from COVID-19 contribute to a high activation value for these cases. At most decision thresholds, these images are examples of true positive classifications. Note the network also correctly underweights the subdiaphragmatic regions and the lower portions of the mediastinum, basing the decision rather on potential lung abnormalities. Mid-level weight is given to large areas of relatively clear lung (green-yellow).

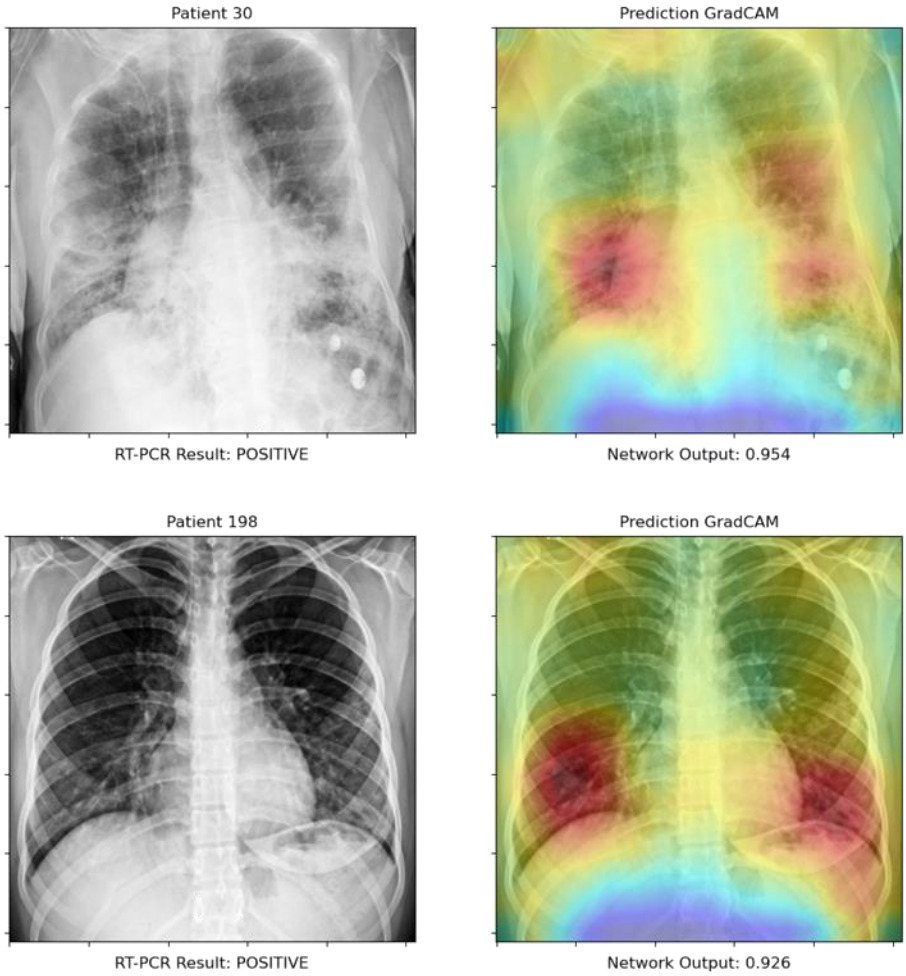

Figure 2. Grad-CAM activation maps for two clinically acquired images of COVID-19 patients with a positive RT-PCR result.

For comparison, Figure 3 shows a typical negative classification pattern for CVX. Here we see much of the peripheral area surrounding the lung is low-weighted in the classification decision. Relatively uniform higher weight is given to the majority of the lung, with the most intense weighting on clear lung portions (red). The detected clear lung regions contributes to a low network output. At most decision thresholds, these two cases are examples of activation patterns for true negative classifications.

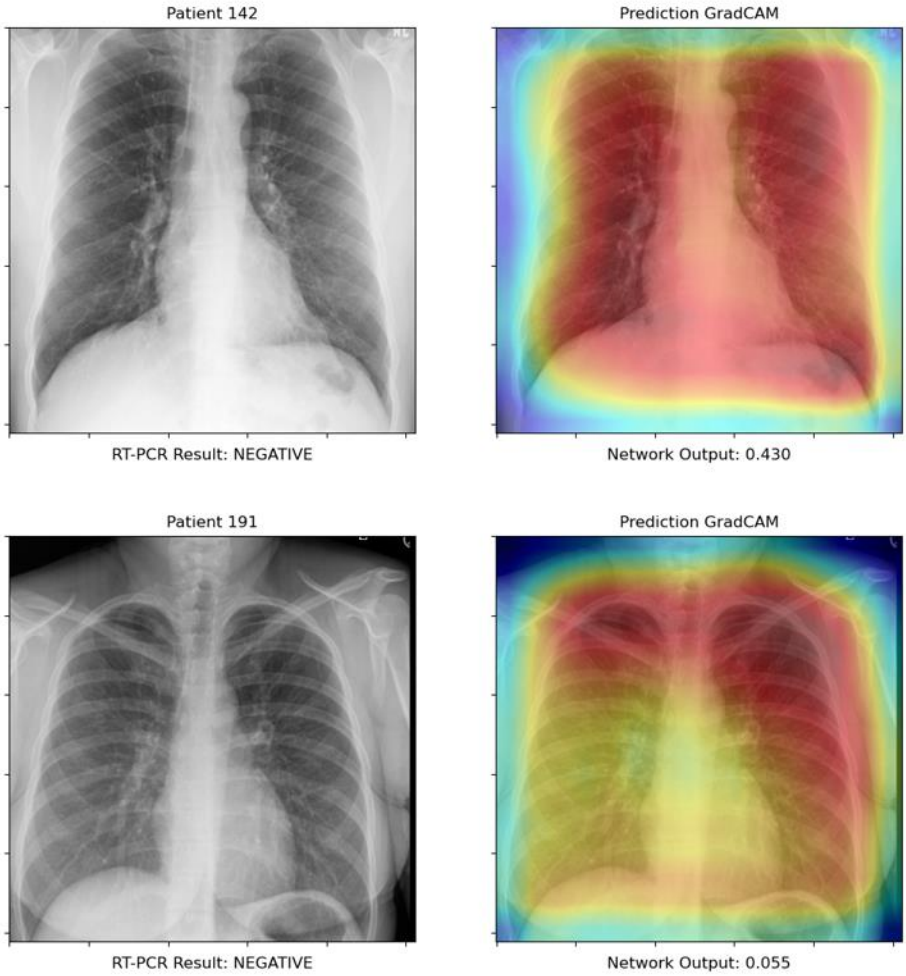

Figure 3. Grad-CAM activation maps for two clinical images of COVID-19 patients with a negative RT-PCR result.

3.2 Application to simulated images

Using both models for inference on the set of 120 simulated radiography images, Figure 4 shows the ROC plots for model predictions compared to the known disease state in these images. We observe classification performance equivalent to the clinical image set when the clinically trained COVID-19 classifier CVX is applied to simulated images. The CVX model generalizes well from the clinically acquired images to the simulated images, with some performance degradation as shown by the diminished AUC of 0.764 (95% CI: 0.680 – 0.848). In contrast, a model fine-tuned on the set of COVID-19 clinical images using a more standard transfer learning approach (STL) loses essentially all its predictive value when applied to the simulated images. The STL model fails to generalize and achieves AUC of 0.524 (95% CI: 0.413 – 0.634). The performance difference on simulated data is clearly distinguishable at the significance level of $\alpha = 0.05$ with p-value of 1.165×10^{-6} as compared with DeLong's method.

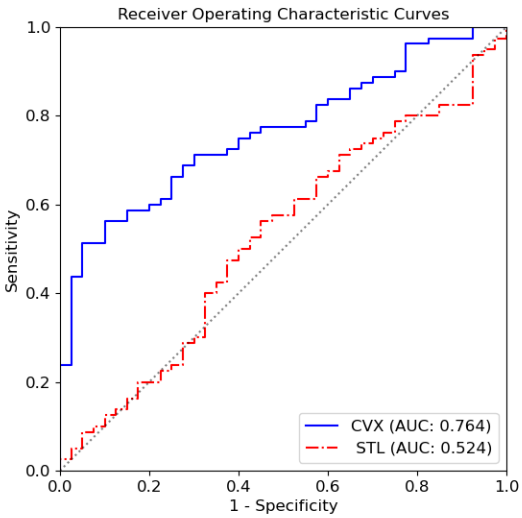

Figure 4. Comparison of receiver operating characteristic curves (ROC) for the COVID-19 X-ray classifier (CVX) to a standard transfer learning approach (STL) on simulated images. The CVX model adequately classifies the simulated images, whereas STL loses its predictive value. The performance difference is a statistically significant at $\alpha = 0.05$ on this set, with p-value of 1.165×10^{-6}.

Simulated radiographs enable us to image the same patient with various COVID-19 findings present and absent. Here we employ Grad-CAM saliency maps to examine if the network is detecting the addition of COVID-19 findings. Figure 5 shows the same habitus imaged twice: in the upper row with no COVID-19 imaging findings, in the lower row with prominent consolidations, particularly in the upper right lung. In the center column of Figure 5, we see that a moderate output of the CVX model in the negative example refines its weighting sharply onto a detected lesion in the positive example, with an equivalent increase in network output activation. The CVX network effectively detects the inserted lesion. By comparison, the STL network has relatively random activations that are not cueing into specific findings, and in fact the network output decreases significantly when a lesion is inserted.

Another example is provided in Figure 6. Here the primary difference is the addition of a subtle finding in the lower right lobe of the lung, just above the right hemidiaphragm and liver. Once again CVX saliency maps indicate a response to this insertion, as the weighting becomes more concentrated around the expected region. We also see the CVX output increase (from 0.188 to 0.354), however this subtle case is unlikely to be positively labeled at most decision thresholds. STL again shows counterproductive behavior, with randomized focal points in both cases and a decreased output score in the positive case compared to the negative case.

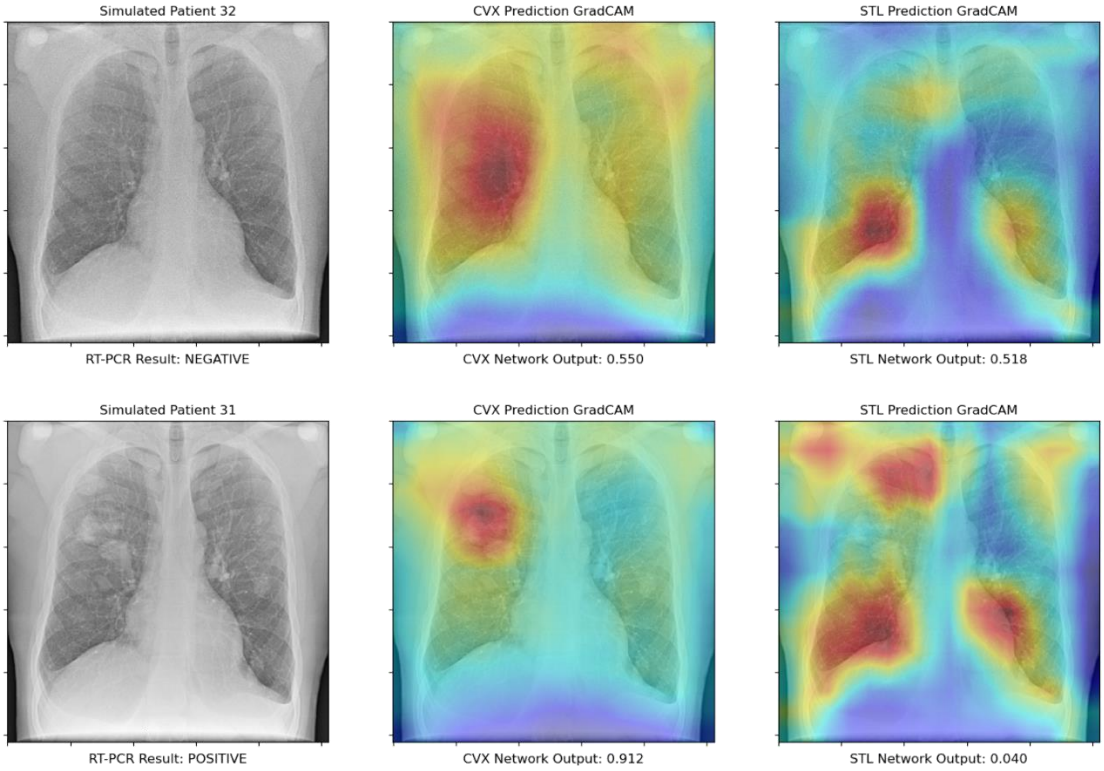

Figure 5. Saliency maps for two simulated patient images based on the same habitus and equivalent imaging conditions. The upper row (Patient 32) is a standard phantom with no COVID-19 imaging features inserted, whereas the lower row (Patient 31) has had several inserted opacities. The CVX network correspondingly increases the weight given to an area with prominent imaging findings, and increases its network output, providing two separable outputs for classification. The STL network has no clear pattern when applied to simulated images and predicts that COVID-19 is less likely when the lesions are inserted.

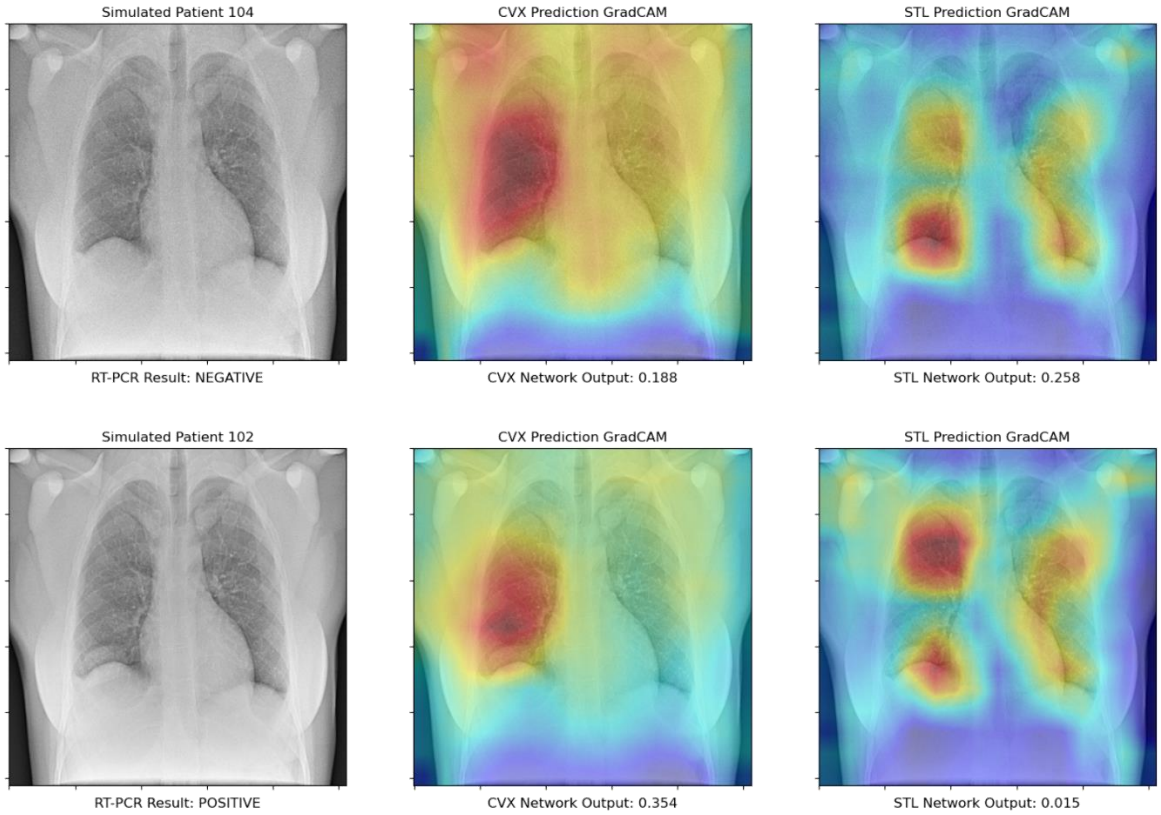

Figure 6. Saliency maps for two simulated patient images based on the same habitus and equivalent imaging conditions. The upper row (Patient 104) is a standard phantom with no COVID-19 imaging features inserted, whereas the lower row (Patient 102) has a subtle finding inserted in the lower right lung. The CVX network focuses the weight given to the appropriate area, however both cases are rated as unlikely for COVID-19. The STL network has consistent but counterproductive patterns that result in a decreased output when the finding is present.

3.3 Stratification by imaging parameters

For the remainder of the evaluation, we apply the CVX network to the simulated images to attempt to isolate the effects of imaging conditions on classifier accuracy. When evaluating results conditioned on exposure in mAs, the ROC curves in Figure 7 have limited resolution due to sample size (30 per exposure level) and sample a range of exposure levels with discretely spaced values. From ROC analysis on these dimensions in isolation, exposure does not consistently account for classifier performance, as the AUC does not correlate with mAs level. At most exposure levels the AUC is consistent at near 0.800. The selected range includes some mild over- and under-exposure of the radiograph for the cases simulated. We infer from these results that the CVX classifier is able to accommodate and 'read through' the exposure level, as performance is invariant to this variable.

Further stratifying results by patient diameter, we start to see possible reasons for an unclear relation between exposure and classification accuracy alone. Figure 8 shows classification results when a hypothetical population-optimal decision threshold is set. We set a threshold of 0.46, where network outputs above this value are predicted to be positive. This threshold value roughly maximizes the F1-score for COVID-19 classification (precision = 0.89, recall = 0.59, F1 = 0.71). A greater proportion of misclassifications are present in smaller diameter patients, with patients under 23cm diameter experiencing 18 of the total 33 false negatives. 11 of those 18 false negatives are in the higher exposure levels of 2 mAs and 4 mAs. Patients above 23cm diameter account for 51 of 81 correct classifications, with all 6 of the observed false

positive results in patients with diameters above 24.3 cm. These results indicate that higher exposures may be detrimental to classification accuracy in some patients, and that a population-optimal decision threshold disproportionately distributes errors within the patient population.

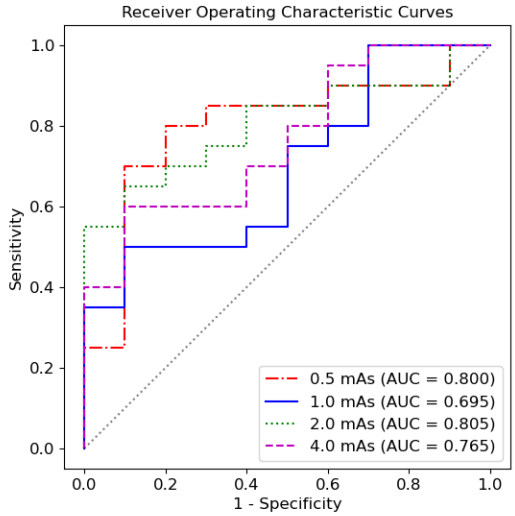

Figure 6. ROC curves for simulated images, stratified by the exposure level in mAs. For each exposure level, the same combination of imaging factors such as habitus and underlying pathology are replicated. We find that varying exposure level does not correlate with classifier performance as measured by AUC.

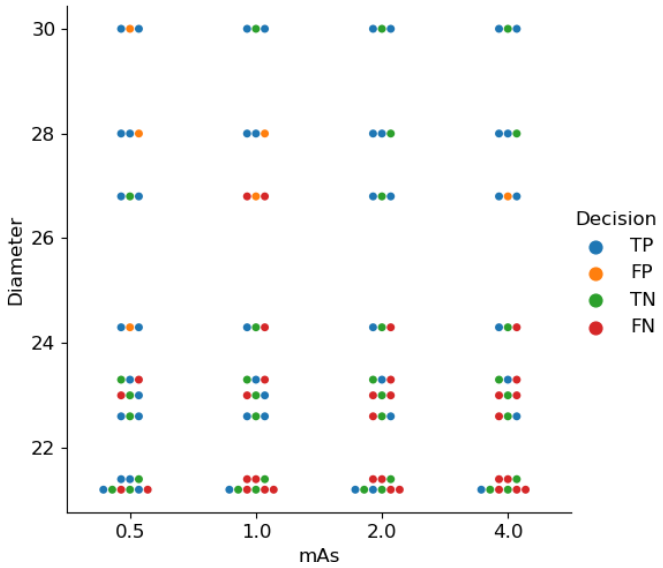

Figure 6. Classification and misclassification decisions at a set decision threshold. The threshold is set to optimize the F1 score in the overall simulated data set. At this decision threshold most misclassifications occur in smaller patients (diameter <23cm), particularly as exposure increases.

4 DISCUSSION

New applications of machine learning in computer-aided diagnostics must contend with the finite collection rate of acceptable data. Some applications, such as the respiratory disease COVID-19, may benefit from existing databases of similar data. In cases where a clear analogous data set is unavailable, simulated data may viably supplement clinical data sets, provided the simulated data is sufficiently realistic or disease features can be readily specified. In our examination, we find that our x-ray classifier generalizes well to simulated images of COVID-19. This is supported by the similarity in both quantitative metrics and activation patterns in the cases examined. In addition to comparable performance in quantitative metrics, the activation maps visually demonstrate that the classifier is cueing into critical features inserted into simulated data in patterns consistent with clinical data. Therefore, it appears from an inference perspective the simulated images are functionally equivalent to clinical data. It remains to be seen if supplementing training data with simulated cases yields an appreciable improvement in algorithm performance. To our knowledge, data augmentation via virtual imaging simulation has not been explored due to limitations in readily producing a large volume of realistic simulated images. Our work in virtual imaging trials looks to expand these capabilities.

The capability to produce viable, functionally equivalent training sets with simulative control over imaging characteristics opens the possibility of a variety of novel experimentation pathways. We explored a subset of imaging parameters and these results imply a complex interaction between imaging settings and classification accuracy. The ability to add or remove lesions in a simulative patient and provide multiple imaging studies allows for true matched pairing experimental designs in evaluating model performance. We can hold imaging factors constant to optimize lesion detection, and conversely optimize the imaging protocol to enhance classifier performance. This preliminary investigation indicates that classification algorithms may be able to adapt to noisy image presentations, as in Figure 6 classifier performance is independent of the exposure level.

A limitation of the current study lies in the small sample of simulated data available. Each simulation combines more deterministic factors such as the choice of initial habitus or imaging system represented, with stochastic factors such as the noise generated in each image produced. We examined only a subset of these variables by including a few variations of habitus, inserted lesion, exposure level, patient diameter (strongly dependent on habitus), a fixed imaging system, and some variation in noise and scattering in each realization of the scatter process represented. Given the large number of variables that influence imaging conditions, the total number of possible imaging conditions considering all permutations of even the factors presented here make exhaustively searching this space computationally prohibitive. In subsequent studies we aim to focus representing the practical range of a handful of key variables, such as exposure level, while further sampling the remaining variables, such as patient size and lesion morphology. A larger more substantive set may more clearly differentiate the influence of these variables on classifier performance in isolation. In turn, classifiers can be further trained on a more substantive set to improve classifier performance for a targeted subset of imaging conditions.

5 CONCLUSIONS

The simulated images appear functionally equivalent to clinical images from the perspective of the classification model in this evaluation. Given that classifier behavior on simulated images is similar to clinical results, the simulated images allow for further investigation into patient-specific optimal imaging protocols. We see that these patient-level factors considerably influence classification accuracy, in combination with imaging parameters such as exposure level.

REFERENCES

[1] Rubin, G. D., Ryerson, C. J., Haramati, L. B., Sverzellati, N., Kanne, J. P., Raoof, S., Schluger, N. W., Volpi, A., Yim, J.-J., Martin, I. B. K., Anderson, D. J., Kong, C., Altes, T., Bush, A., Desai, S. R., Goldin, J., Goo, J. M., Humbert, M., Inoue, Y., et al., "The Role of Chest Imaging in Patient Management during the COVID-19 Pandemic: A Multinational Consensus Statement from the Fleischner Society," Radiology, 201365 (2020).

[2] Desai, S., Baghal, A., Wongsurawat, T., Al-Shukri, S., Gates, K., Farmer, P., Rutherford, M., Blake, G. D., Nolan, T., Powell, T., Sexton, K., Bennett, W. and Prior, F., "Chest Imaging with Clinical and Genomic Correlates Representing a Rural COVID-19 Positive Population" (2020).

[3] "NIH harnesses AI for COVID-19 diagnosis, treatment, and monitoring.", <https://www.nibib.nih.gov/news-events/newsroom/nih-harnesses-ai-covid-19-diagnosis-treatment-and-monitoring> (12 August 2020).

[4] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M. and Summers, R. M., "ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases," 2017 IEEE Conf. Comput. Vis. Pattern Recognit. CVPR, 3462–3471 (2017).
[5] Baltruschat, I. M., Nickisch, H., Grass, M., Knopp, T. and Saalbach, A., "Comparison of Deep Learning Approaches for Multi-Label Chest X-Ray Classification," 1, Sci. Rep. **9**(1), 1–10 (2019).
[6] "National Lung Screening Trial - National Lung Screening Trial (NLST) Public Access - Cancer Imaging Archive Wiki.", <https://wiki.cancerimagingarchive.net/display/NLST/National+Lung+Screening+Trial#cca39708a07d4cfa8e4685bcf6002070> (12 August 2020).
[7] Armato, S. G., McLennan, G., Bidaut, L., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., Zhao, B., Aberle, D. R., Henschke, C. I., Hoffman, E. A., Kazerooni, E. A., MacMahon, H., van Beek, E. J. R., Yankelevitz, D., Biancardi, A. M., Bland, P. H., Brown, M. S., Engelmann, R. M., Laderach, G. E., et al., "The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A Completed Reference Database of Lung Nodules on CT Scans: The LIDC/IDRI thoracic CT database of lung nodules," Med. Phys. **38**(2), 915–931 (2011).
[8] Gohagan, J. K., Prorok, P. C., Hayes, R. B., Kramer, B. S., and Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial Project Team., "The Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial of the National Cancer Institute: history, organization, and status," Control. Clin. Trials **21**(6 Suppl), 251S-272S (2000).
[9] Fricks, R., Ria, F., Chalian, H., Khoshpouri, P., Abadi, E., Bianchi, L., Segars, W. P. and Samei, E., "Deep Learning Classification of COVID-19 in Chest Radiographs: Performance and Influence of Supplemental Training" (2020).
[10] "RSNA Pneumonia Detection Challenge.", <https://kaggle.com/c/rsna-pneumonia-detection-challenge> (19 June 2020).
[11] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. and Batra, D., "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," 2017 IEEE Int. Conf. Comput. Vis. ICCV, 618–626 (2017).
[12] Huang, G., Liu, Z., van der Maaten, L. and Weinberger, K. Q., "Densely Connected Convolutional Networks," ArXiv160806993 Cs (2018).
[13] Deng, J., Dong, W., Socher, R., Li, L.-J., Kai Li and Li Fei-Fei., "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conf. Comput. Vis. Pattern Recognit., 248–255 (2009).
[14] Bengio, Y., Louradour, J., Collobert, R. and Weston, J., "Curriculum learning," Proc. 26th Annu. Int. Conf. Mach. Learn., 41–48, Association for Computing Machinery, Montreal, Quebec, Canada (2009).
[15] Sun, X. and Xu, W., "Fast Implementation of DeLong's Algorithm for Comparing the Areas Under Correlated Receiver Operating Characteristic Curves," IEEE Signal Process. Lett. **21**(11), 1389–1393 (2014).
[16] DeLong, E. R., DeLong, D. M. and Clarke-Pearson, D. L., "Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach," Biometrics **44**(3), 837–845 (1988).
[17] Abadi, E., Segars, W. P., Chalian, H. and Samei, E., "Virtual Imaging Trials for Coronavirus Disease (COVID-19)," Am. J. Roentgenol. **216**(2), 1–7 (2021).
[18] Abadi, E., Harrawood, B., Sharma, S., Kapadia, A., Segars, W. P. and Samei, E., "DukeSim: A Realistic, Rapid, and Scanner-Specific Simulation Framework in Computed Tomography," IEEE Trans. Med. Imaging **38**(6), 1457–1465 (2019).
[19] Sharma, S., Kapadia, A., Fu, W., Abadi, E., Segars, W. P. and Samei, E., "A real-time Monte Carlo tool for individualized dose estimations in clinical CT," Phys. Med. Biol. **64**(21), 215020 (2019).