10.1148/radiol.2020202944

Appendix E1.

Bias Mitigation

In image classification tasks, machine learning methods use image intensity values to extract digital image features and then use these image features to compute, for example, the COVID-19 probability score of an input CXR. Therefore, the actual digital values in each CXR image determine the final machine learning classification decision of the input CXR. However, different vendors of CXR imaging systems use different proprietary postprocessing algorithms to process digital CXRs (ie, each vendor will adjust their final digital values differently for desired image contrast for interpretation). Further, many hospitals and clinics often use multiple CXR imaging systems from different vendors. Additionally, different clinics and different technologists may choose different imaging parameters such as x-ray tube potential (kVp values) and x-ray exposure levels (mAs values) to acquire the CXR. As a result, similar pneumonia findings may have very different digital image representations in retrospectively collected digital CXRs. Without taking these variables into account, machine learning algorithms may produce biased results.

Image Preprocessing

The DICOM files were resized to 1024×1024 and saved as 8-bit PNG grayscale images. The image intensity value was adjusted based on the window level and window width attributes in the DICOM file. Contrast inversion is applied for images with DICOM attribute MONOCHROME1. See Figure E1 for the flowchart of the image preprocessing step.

Before being fed into the network, PNG images were further downscaled to 224×224, converted to red (R)-green (G)-blue (B) images and normalized based on the mean and standard deviation of images in the ImageNet dataset:

$$R = (R - 0.485) / 0.229$$

$$G = (G - 0.456) / 0.224$$

$$B = (B - 0.406) / 0.225$$

Network Architecture and Training

The DenseNet-121 (1) architecture with 50 convolutional neural network (CNN) layers was used as the image feature extraction module of CV19-Net. Followed by the last convolutional layer of DenseNet-121 (layer 120) is a fully connected (FC) classifier with a Softmax activation function to combine the extracted feature vector for the final predicted probability score.

A three-stage transfer learning process was used in model training (see Fig E2):

1. Stage one: The image feature extraction module was trained on ImageNet images with 14 million images to differentiate between 1,000 image classes.

2. Stage two: The network was initialized with weights trained in stage one and was further trained using the NIH data set with 112,120 chest x-ray images from 30,805 unique

patients to classify chest x-ray images into 14 different disease classes. A similar design was used in CheXNet (2). This step allows the network to learn CXR-specific image features.

3. Stage three: The network was initialized using weights obtained from stage two and trained using our training dataset consists of 5,236 CXRs (2,582 CXRs from the COVID-19 cohort and 2,654 CXRs from the non-COVID-19 pneumonia cohort) to train the network to perform the final binary classification: COVID-19 and non-COVID-19 pneumonia classification.

CV19-Net (Fig E3) was developed using the PyTorch framework. The network was trained to minimize the binary cross entropy loss. Adam optimizer was used with an initial learning rate = 6.0×10^{-5} for all convolutional layers and 1.0×10^{-4} for the FC classifier. The minibatch size was empirically selected to be 50. Data augmentation techniques including rotation (30-degree range) and horizontal flipping were used. To prevent model overfitting, an early stopping strategy was adopted by monitoring the training loss on the validation set. The validation set was randomly sampled from the total training dataset (25% of the training samples). The model with the lowest validation loss was taken as the final model for prediction. To reduce fluctuations of prediction results, the well-known ensemble averaging technique common in machine learning was introduced in this work. The prediction scores of input images from $n = 20$ individually trained networks with identical training parameters, but different random seeds in model initialization and different randomly sampled validation sets. A quadratic mean of the prediction probability scores was taken to generate the final predication probability score:

$$S = \left[\frac{1}{N} \sum_{i=1}^{N} S^2(i) \right]^{1/2} \quad (n = 20).$$

This final probability score was compared with the selected threshold values in decision making to perform binary classification.

Class Activation Maps

To help visualize which part of the input images contributed most to CV19-Net's decision used to produce the final probability score, a heat map employing the gradient-weighted Class Activation Mapping (Grad-CAM) (3) was used to highlight those key image pixels in the CXR image primarily responsible for COVID-19 pneumonia. The paired heatmap of COVID-19 image features and the original CXR images are presented in Figure E4 to help aid human eyes to identify the key morphologic and contextual features in CXR images.

Additional Statistical Analyses

(a) Test Performance Difference on Men and Women

The following statistical hypothesis testing was performed:

$$H_0 : AUC(male) = AUC(female) \quad \text{versus}$$

$$H_1 : AUC(male) \neq AUC(female)$$

P value method was used in hypothesis testing with a rejection *P* value of .05. Result shows *P* = .17, therefore no statistically significant difference between two groups.

(b) Test Performance Difference on Patients of Different Age Groups

The following statistical hypothesis testing was performed:

$$H_0 : AUC\left(Group - i\right) = AUC\left(Group - j\right) \text{ versus}$$

$$H_1 : AUC\left(Group - i\right) \neq AUC\left(Group - j\right)$$

Results are shown in Table E1.

References

1. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, July 21–26, 2017. Piscataway, NJ: IEEE, 2017.

2. Rajpurkar P, Irvin J, Ball RL, et al. Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLoS Med 2018;15(11):e1002686 https://doi.org/10.1371/journal.pmed.1002686.

3. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In: 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, October 22–29, 2017. Piscataway, NJ: IEEE, 2017.

4. Pereira RM, Bertolini D, Teixeira LO, Silla CN Jr, Costa YMG. COVID-19 identification in chest X-ray images on flat and hierarchical classification scenarios. Comput Methods Programs Biomed 2020;194:105532.

5. Cohen JP, Morrison P, Dao L. COVID-19 Image Data Collection. arXiv preprint arXiv:2003.11597. https://arxiv.org/abs/2003.11597. Submitted March 25, 2020. Accessed DATE.

6. Rahimzadeh M, Attar A. A modified deep convolutional neural network for detecting COVID-19 and pneumonia from chest X-ray images based on the concatenation of Xception and ResNet50V2. Inform Med Unlocked 2020;19:100360.

7. Zokaeinikoo M, Kazemian P, Mitra P, Kumara S. AIDCOV: An Interpretable Artificial Intelligence Model for Detection of COVID-19 from Chest Radiography Images. medRxiv. https://doi.org/10.1101/2020.05.24.20111922. Posted June 29, 2020. Accessed DATE.

8. Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Rajendra Acharya U. Automated detection of COVID-19 cases using deep neural networks with X-ray images. Comput Biol Med 2020;121:103792.

9. Jamil M, Hussain I. Automatic Detection of COVID-19 Infection from Chest X-ray using Deep Learning. medRxiv. https://doi.org/10.1101/2020.05.10.20097063. Posted May 14, 2020. Accessed DATE.

10. Narin A, Kaya C, Pamuk Z. Automatic Detection of Coronavirus Disease (COVID-19) Using X-ray Images and Deep Convolutional Neural Networks. arXiv preprint arXiv:2003.10849. https://arxiv.org/abs/2003.10849. Submitted March 24, 2020. Accessed DATE.

11. Gil D, Díaz-Chito K, Sánchez C, Hernández-Sabaté A. Early Screening of SARS-CoV-2 by Intelligent Analysis of X-Ray Images. arXiv preprint arXiv:2005.13928. https://arxiv.org/abs/2005.13928. Submitted May 28, 2020. Accessed DATE.

12. Horry MJ, Paul M, Ulhaq A, Pradhan B, Saha M, Shukla N. X-Ray Image based COVID-19 Detection using Pre-trained Deep Learning Models.. Submitted April 20, 2020. Accessed DATE.

13. Khan AI, Shah JL, Bhat MM. CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. Comput Methods Programs Biomed 2020;196:105581.

14. El Asnaoui K, Chawki Y. Using X-ray images and deep learning for automated detection of coronavirus disease. J Biomol Struct Dyn 2020;9:1–12.

15. Afshar P, Heidarian S, Naderkhani F, Oikonomou A, Plataniotis KN, Mohammadi A. COVID-CAPS: A Capsule Network-based Framework for Identification of COVID-19 cases from X-ray Images. arXiv preprint arXiv:2004.02696. https://arxiv.org/abs/2004.02696. Submitted April 6, 2020. Accessed DATE.

16. Karim MR, Döhmen T, Rebholz-Schuhmann D, Decker S, Cochez M, Beyan O. DeepCOVIDExplainer: Explainable COVID-19 Diagnosis Based on Chest X-ray Images. arXiv preprint arXiv:2004.04582. https://arxiv.org/abs/2004.04582. Submitted April 9, 2020. Accessed DATE.

17. Oh Y, Park S, Ye JC. Deep learning covid-19 features on cxr using limited training data sets. IEEE Trans Med Imaging 2020;39(8):2688–2700 https://doi.org/10.1109/TMI.2020.2993291.

18. Wang L, Wong A. COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images. arXiv preprint arXiv:2003.09871. https://arxiv.org/abs/2003.09871. Submitted March 22, 2020. Accessed DATE.

19. Luz E, Silva PL, Silva R, Silva L, Moreira G, Menotti D. Towards an Effective and Efficient Deep Learning Model for COVID-19 Patterns Detection in X-ray Images. arXiv preprint arXiv:2004.05717. https://arxiv.org/abs/2004.05717. Submitted April 12, 2020. Accessed DATE.

20. Ucar F, Korkmaz D. COVIDiagnosis-Net: Deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images. Med Hypotheses 2020;140:109761.

21. Farooq M, Hafeez A. COVID-ResNet: A Deep Learning Framework for Screening of COVID19 from Radiographs. arXiv preprint arXiv:2003.14395. https://arxiv.org/abs/2003.14395. Submitted March 31, 2020. Accessed DATE.

22. Shibly KH, Dey SK, Islam MTU, Rahman MM. COVID Faster R-CNN: A Novel Framework to Diagnose Novel Coronavirus Disease (COVID-19) in X-Ray Images. medRxiv. https://doi.org/10.1101/2020.05.14.20101873. Posted DATE. Accessed DATE.

23. Majeed T, Rashid R, Ali D, Asaad A. Problems of Deploying CNN Transfer Learning to Detect COVID-19 from Chest X-rays. medRxiv.. Posted June 7, 2020. Accessed DATE.

24. Zhang Y, Niu S, Qiu Z, et al. COVID-DA: Deep Domain Adaptation from Typical Pneumonia to COVID-19. arXiv preprint arXiv:2005.01577. https://arxiv.org/abs/2005.01577. Submitted April 30, 2020. Accessed DATE.

25. Kumar R, Arora R, Bansal V, et al. Accurate Prediction of COVID-19 using Chest X-Ray Images through Deep Feature Learning model with SMOTE and Machine Learning Classifiers. medRxiv.. Submitted April 17, 2020. Accessed DATE.

26. Tahir A, Qiblawey Y, Khandakar A, et al. Coronavirus: Comparing COVID-19, SARS and MERS in the eyes of AI. arXiv preprint arXiv:2005.11524. https://arxiv.org/abs/2005.11524. Submitted May 23, 2020. Accessed DATE.

27. Yeh CF, Cheng HT, Wei A, et al. A Cascaded Learning Strategy for Robust COVID-19 Pneumonia Chest X-Ray Screening. arXiv preprint arXiv:2004.12786. https://arxiv.org/abs/2004.12786. Submitted April 24, 2020. Accessed DATE.

28. Schwab P, Schütte AD, Dietz B, Bauer S. predCOVID-19: A Systematic Study of Clinical Predictive Models for Coronavirus Disease 2019. arXiv preprint arXiv:2005.08302. https://arxiv.org/abs/2005.08302. Submitted May 17, 2020. Accessed DATE.

29. Murphy K, Smits H, Knoops AJG, et al. COVID-19 on Chest Radiographs: A Multireader Evaluation of an Artificial Intelligence System. Radiology 2020;296(3):E166–E172.

Table E1: Paired AUC difference between different age groups

Age group AUC difference, *P* value					
	18–30	30–45	45–60	60–75	75–90
18–30		0.31	0.02	0.002	<0.001
30–45			0.13	0.01	<0.001
45–60				0.25	0.002
60–75					0.02
75–90					

Table E2: Related works

Reference	Number of positive CXRs in training/validation	Number of positive CXRs in testing	Data type
Pereira et al (4)	63	27	Cohen (5)
Rahimzadeh et al (6)	149	31	Cohen
Zokaeinikoo et al (7)	267		Cohen
Ozturk et al (8)	127		Cohen
Kishore et al (9)	150		Cohen
Narin et al (10)	269		Cohen
Gil et al (11)	288		Cohen
Horry et al (12)	100		Cohen
Khan et al (13)	284		Cohen
Elasnaoui et al (14)	231		Cohen
Afshar et al (15)	Not clear		Cohen
Karim et al (16)	149	31	Cohen
Oh et al (17)	144	36	Cohen
Wang et al (18)	358		Wang
Luz et al (19)	152	31	Wang
Ucar et al (20)	66	10	Wang

Farooq et al (21)	68		Wang
Shibly et al (22)	232	51	Wang
Majeed et al (23)	111	73	Cohen, Kaggle
Zhang et al (24)	258	60	Cohen, Kaggle
Kumar et al (25)	42	20	SIRM
Tahir et al (26)	338	85	Cohen, SIRM, Radiopaedia
Yeh et al (27)	415	95	Local hospital
Schwab et al (28)	391	167	Local hospital
Murphy et al (29)	512	468	Local hospital

SIRM: https://www.sirm.org/category/senza-categoria/covid-19., Cohen: https://github.com/ieee8023/covid-chestxray-dataset., Kaggle: https://www.kaggle.com/andrewmvd/covid19-X-rays., Wang: https://github.com/lindawangg/COVID-Net., Radiopaedia: https://radiopaedia.org/playlists/25975?.