# Deep-learning convolutional neural networks with transfer learning accurately classify COVID19 lung infection on portable chest radiographs

**Shreeja Kikkisetti** [1], **Jocelyn Zhu** [1], **Beiyi Shen** [2], **Haifang Li** [2], **Tim Duong** [Corresp. 1]

[1] Radiology, Montefiore Medical Center and Albert Einstein College of Medicine, Bronx, New York, United States

[2] Radiology, State University of New York at Stony Brook, Stony Brook, New York, United States

Corresponding Author: Tim Duong
Email address: tim.duong@stonybrook.edu

Portable chest x-ray (pCXR) has become an indispensable tool in the management of Coronavirus Disease 2019 (COVID-19) lung infection. This study employed deep-learning convolutional neural networks to classify COVID-19 lung infections on pCXR from normal and related lung infections to potentially enable more timely and accurate diagnosis. This retrospect study employed deep-learning convolutional neural network (CNN) with transfer learning to classify based on pCXRs COVID-19 pneumonia (N=455) on pCXR from normal (N=532), bacterial pneumonia (N=492), and non-COVID viral pneumonia (N=552). The data was randomly split into 75% training and 25% testing, randomly. A five-fold cross-validation was used for the testing set separately. Performance was evaluated using receiver-operating curve analysis. Comparison was made with CNN operated on the whole pCXR and segmented lungs. CNN accurately classified COVID-19 pCXR from those of normal, bacterial pneumonia, and non-COVID-19 viral pneumonia patients in a multiclass model. The overall sensitivity, specificity, accuracy, and AUC were 0.79, 0.93, and 0.79, 0.85 respectively (whole pCXR), and were 0.91, 0.93, 0.88, and 0.89 (CXR of segmented lung). The performance was generally better using segmented lungs. Heatmaps showed that CNN accurately localized areas of hazy appearance, ground glass opacity and/or consolidation on the pCXR. Deep-learning convolutional neural network with transfer learning accurately classifies COVID-19 on portable chest x-ray against normal, bacterial pneumonia or non-COVID viral pneumonia. This approach has the potential to help radiologists and frontline physicians by providing more timely and accurate diagnosis.

1 **Deep-learning convolutional neural networks with transfer learning**

2 **accurately classify COVID19 lung infection on portable chest**

3 **radiographs**

5 Shreeja Kikkisetti[1], Jocelyn Zhu[1], Beiyi Shen[2], Haifang Li[2], Tim Q Duong[1]

6 [1] Radiology, Montefiore Medical Center and Albert Einstein College of Medicine, Bronx,
7 New York, United States

8 [2] Radiology, State University of New York at Stony Brook, Stony Brook, New York, United
9 States

11
12
13 Correspondence: Tim Duong, Ph.D., Radiology, Montefiore Medical Center and Albert Einstein
14 College of Medicine, 111 E 210th St, Bronx, NY 10467 Tim.duong@einsteinmed.org, Tel: 718
15 920 6268

17 Funding: none

18 Manuscript Type: Original Investigation

19 **Abbreviations**
20
21     COVID-19    Coronavirus Disease 2019
22     CXR    chest x-ray
23     CNN    convolutional neural networks
24     AI    artificial intelligence

25 **Acknowledgement**

26      none

27   Word Count for Text: 2700

29   **ABSTRACT**

30   Portable chest x-ray (pCXR) has become an indispensable tool in the management of Coronavirus

31   Disease 2019 (COVID-19) lung infection. This study employed deep-learning convolutional

32   neural networks to classify COVID-19 lung infections on pCXR from normal and related lung

33   infections to potentially enable more timely and accurate diagnosis. This retrospect study

34   employed deep-learning convolutional neural network (CNN) with transfer learning to classify

35   based on pCXRs COVID-19 pneumonia (N=455) on pCXR from normal (N=532), bacterial

36   pneumonia (N=492), and non-COVID viral pneumonia (N=552). The data was randomly split into

37   75% training and 25% testing, randomly. A five-fold cross-validation was used for the testing set

38   separately. Performance was evaluated using receiver-operating curve analysis. Comparison was

39   made with CNN operated on the whole pCXR and segmented lungs. CNN accurately classified

40   COVID-19 pCXR from those of normal, bacterial pneumonia, and non-COVID-19 viral

41   pneumonia patients in a multiclass model. The overall sensitivity, specificity, accuracy, and AUC

42   were 0.79, 0.93, and 0.79, 0.85 respectively (whole pCXR), and were 0.91, 0.93, 0.88, and 0.89

43   (CXR of segmented lung). The performance was generally better using segmented lungs.

44   Heatmaps showed that CNN accurately localized areas of hazy appearance, ground glass opacity

45   and/or consolidation on the pCXR. Deep-learning convolutional neural network with transfer

46   learning accurately classifies COVID-19 on portable chest x-ray against normal, bacterial

47    pneumonia or non-COVID viral pneumonia. This approach has the potential to help radiologists

48    and frontline physicians by providing more timely and accurate diagnosis.

49

50    **Key words:**

51      coronavirus; machine learning; chest X-ray; computed tomography; lung infection

52

53    **Introduction**

54          Coronavirus Disease 2019 (COVID-19) is a highly infectious disease that causes severe

55    respiratory illness (Hui et al. 2020; Lu et al. 2020). It was first reported in Wuhan, China in

56    December 2019 (Li et al. 2020c) and was declared a pandemic on Mar 11, 2020 . The first

57    confirmed case of coronavirus disease 2019 (COVID-19) in the United States was reported from

58    Washington State on January 31, 2020 (Holshue et al. 2020). Soon after, Washington, California

59    and New York reported outbreaks. COVID-19 has already infected 10 million, killed more than

60    0.5 million people, and the United States has become the worst-affected country, with more than

61    2.4 million diagnosed cases and at least 122,796 deaths (https://coronavirus.jhu.edu, assessed Jun

62    28, 2020). There are recent spikes of COVID-19 infection cases across many states and around the

63    world and there will likely be second waves and recurrence.

64          A definitive test of COVID-19 infection is the reverse transcription polymerase chain

65    reaction (RT-PCR) of a nasopharyngeal or oropharyngeal swab specimen (Tang et al. 2020; Wang

66    et al. 2020). Although RT-PCR has high specificity, it has low sensitivity, high false negative rate,

67    and long turn-around time (Tang et al. 2020; Wang et al. 2020) (currently ~4 days although

68    improvement  and other tests are becoming available (CDC). By contrast, portable chest X-rays

69    (pCXR) is convenient to perform, has a fast turnaround, and is well suited for imaging contagious

70    patients and longitudinal monitoring of critically ill patients in the intensive care units because the

71    equipment can be readily disinfected, preventing cross-infection. pCXR of COVID-19 infection

72    has certain unique characteristics, such as predominance of bilateral, peripheral, and low lobes

73    involvement, with ground-glass opacities with or without airspace consolidations as the disease

74    progresses. These characteristics generally differ from other lung pathologies, such as bacterial

75    pneumonia or other viral (non-COVID-19) lung infection. Based on CXR and laboratory findings,

76    clinicians might start patients on empirical treatment before the RT-PCR results become available

77    or even if the RT-PCR come back negative due to high false negative rate of RT-PCR. Early

78    treatment in COVID-19 patients is associated with better clinical outcomes. Similarly, computed

79    tomography (CT), which offers relatively more detailed features (such as subtle ground-glass

80    opacity (Li et al. 2020b; Xu et al. 2020)), has also been used in the context of COVID-19. However,

81    CT suite and equipment are more challenging to disinfect, and thus it is much less suitable for

82    examining patients suspected of or confirmed with contagious diseases in general and COVID-19

83    in particular. Longitudinal CT monitoring of critically ill patients in the intensive care units is also

84    challenging. In short, pCXR has become an indispensable imaging tool in the management of

85    COVID-19 infection, is often one of the first examinations a patient suspected of COVID-19

86    infection receives in the emergency room, and ideally used for longitudinal monitoring of critically

87    ill patients in the intensive care units.

88          The usage of pCXR under the COVID-19 pandemic circumstances is unusual in many

89    aspects. For instance, pCXR is preferred as it can be used at the bedside without moving the

90    patients, but the imaging quality is not as good as conventional CXR. In addition, COVID-19

91    patients may not be able to take full inspirations during the examination, obscuring possible

92    pathology, especially in the lower lung fields. Many sicker patients may be positioned on the side

93   which compromises imaging quality. Thus, pCXR data under the COVID-19 pandemic

94   circumstances are suboptimal and, thus, may be more challenging to interpret. Moreover, pCXR

95   is increasingly read by non-chest radiologists in some hospitals due to increasing demands,

96   resulting in reduced accuracy and efficiency.

97      pCXR images contain important clinical features that could be easily missed by the naked

98   eyes. Computer-aided methods can improve efficiency and accuracy of pCXR interpretations,

99   which in turn provides more timely and relevant information to frontline physicians. Deep-learning

100  artificial intelligence (AI) is increasingly used to analyze diagnostic images (Ehteshami Bejnordi

101  et al. 2018; LeCun et al. 2015). AI has the potential to facilitate disease diagnosis, staging of

102  disease severity and longitudinal monitoring of disease progression.

103     One common machine-learning algorithm is the convolutional neural network (CNN)

104  (Krizhevsky et al. 2012), which takes an input image, learns important features in the image such

105  as size or intensity, and saves these parameters as weights and bias to differentiate types of images

106  (Song et al. 2020). CNN architecture is ideally suited for analyzing images. Moreover, many of

107  the machine learning algorithms are trained to solve specific tasks, where models need to be rebuilt

108  from scratch if the feature changes. Transfer learning overcomes such drawback by utilizing

109  knowledge acquired for one task to solve related ones. Transfer learning is useful when dealing

110  with small sample size data because the pre-trained weights improve efficiency and performance

111  (Byra et al. 2018; Samala et al. 2017).

112     Many artificial intelligence (AI) algorithms based on deep-learning convolutional neural

113  networks have been deployed for pCXR applications (Harris et al. 2019; Heo et al. 2019; Mekov

114  et al. 2020) and these algorithms can be readily repurposed for COVID-19 pandemic

115  circumstances. While there are already many papers describing prevalence and radiographic

116    features on pCXR of COVID-19 lung infection (see reviews (Bao et al. 2020)), there is a few AI

117    papers (Apostolopoulos & Mpesiana 2020; Cohen et al. 2020a; Elaziz et al. 2020; Hurt et al. 2020;

118    Murphy et al. 2020; Ozturk et al. 2020; Pereira et al. 2020; Zhu et al. 2020a) to classify CXRs of

119    COVID-19 patients from CXR of normals or related lung infections. The full potential of AI

120    applications of pCXR under COVID-19 pandemic circumstances is not yet fully realized.

121          The goal of this pilot study is to employ deep-learning convolutional neural networks to

122    classify normal, bacterial infection, and non-COVID-19 viral infection (such as influenza) against

123    COVID-19 infection on pCXR. The performance was evaluated using receiver-operating curve

124    (ROC) analysis. Heatmaps were also generated to visualize and assessment the performance of the

125    AI algorithm.

126

## Materials and Methods

127

128

129    ***Data sources:*** This retrospective study used publicly available pCXR of i) COVID-19 infection,

130    ii) non-COVID-19 viral infection, iii) bacterial pneumonia, and iv) normal subjects. The COVID-

131    19 pCXR were downloaded from on May 27th, 2020 (Cohen et al. 2020b). The original download

132    contained 673 CT or pCXR images of COVID-19, SARS, acute respiratory distress syndromes,

133    pneumocystis, streptococcus, legionella, Chlamydophila, E Coli, Klebsiella, lipoid, Varicella, and

134    influenza. The labels for the data came from a metadata file associated with the open dataset. The

135    final sample size for COVID-19 patients was 455 pCXR from 197 patients. We recognized that

136    this dataset was a public, community-driven dataset and there are potential selection biases. A

137    radiologist (B.S.) evaluated all images for quality and relevance and each case was COVID-19

138    positive based on available data. As a result of this evaluation, a few images that were deemed to

139    be of poor quality, were excluded.

140        The other datasets were taken from the established Kaggle chest X-ray image (pneumonia)

141    dataset   (https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia).   Although  the

142    Kaggle database has a large sample size, we randomly selected a sample size comparable to that

143    of COVID-19 to avoid asymmetric sample size bias that could skew sensitivity and specificity.

144    The sample sizes chosen for bacterial pneumonia, non-COVID-19 viral pneumonia, and normal

145    pCXR were 492, 552 and 532 patients, respectively. Similarly, a chest radiologist evaluated all

146    images for quality.

147

148        *CNN*: A CNN, a type of neural network, is ideally suited for analyzing images. In a standard

149    CNN model, a filter (window) travels over each region of an image and looks for different features

150    such as edges, colorations, patches, and more in order to classify an image into a certain category.

151    Our CNN architecture was based on VGG16 (**Figure 1**), a convolutional neural network

152    (Simonyan & A. 2014), architecture was utilized for computation efficiency and ease to

153    implement, for immediate translation potential. Our VGG16 architecture had 13 convolution layers

154    that each run a series of filters over the image to extract important features. The VGG16 model

155    was used because it was pretrained on the ImageNet database and properly employs transfer

156    learning which makes the training process efficient. In other words, instead of having to learn all

157    the relationships in an image from scratch, the model is already familiar with that when transfer

158    learning is employed. The data was normalized first by transforming all files into RGB images and

159    resizing them into 224x224 pixels to make them compatible with the VGG16 framework. Next,

160    the images were one-hot-encoded and randomly split into 75% training and 25% testing. One hot

161   encoding means to turn all the categorical labels into numerical values containing zeroes and ones

162   to make it much easier for the computer to read. VGG16 implements 13 convolutional layers: 5

163   Max Pooling layers and 3 Dense layers which sum up to 21 layers and 16 weight layers (Ren et al.

164   2020). Conv 1 has 64 filters while Conv 2 has 128 filters, Conv 3 has 256 filters while Conv 4 and

165   Conv 5 have 512 filters. The first two layers have 2 sublayers while the 4th and 5th layers have 3

166   sublayers. A sublayer is another layer within a convolutional layer that further filters images and

167   passes information down to the next sublayers. The information collected from all the sublayers is

168   compiled and sent to the next layer to make a cohesive prediction. A max-pooling layer was used

169   after each step in the model to down sample the input and identify its important features based on

170   the methods described in (Ren et al. 2020). A max-pooling layer reduces the dimensionality of the

171   image and allows for assumptions to be made about features contained in the sub regions (Ren et

172   al. 2020). All convolutional layers used rectified linear units (ReLUs) as an activation function

173   because it adds a small number of learnable parameters (Ren et al. 2020). Three fully connected

174   layers were used, each having 4096 nodes. Fully connected layers compose some of the last few

175   layers in a model and connect all the inputs from each layer to the activation unit of the next layer

176   (Ren et al. 2020). Dropout layers were used, along with the Softmax function, to prevent

177   overfitting. Dropout layers work by randomly setting the edges of hidden neurons to 0 at each

178   update of the training phase. The softmax function turns all the scores from the images into a

179   normalized probability distribution, which helps make the final prediction (Ren et al. 2020). For

180   data analysis, batch sizes of 32 were used to limit computational expense and trained for 50 epochs.

181   Epochs can be thought of as iterations. Several optimizers were tested and Adams optimization

182   function was found to yield the lowest validation loss. The learning rate was lowered from the

183   recommended 0.01 to 0.001 to prevent overshooting the global minimum loss. Categorical cross

184   entropy was used as a loss function since the loss value decreases as the predicted probability

185   converges to the actual label.

186

187         CNN analysis was performed on the whole pCXR as well as virtually segmented lungs.

188   Lung segmentation was performed using a CNN architecture with 22 convolutional layers, 4 max-

189   pooling layers, and 4 merged layers for connectivity. A ReLu activation function was used with

190   the Keras library. The output consisted of a mask of the segmented lungs. The segmented lungs

191   were then fed into the CNN model for the Covid19 classification. This model was trained on the

192   Montgomery dataset and achieved an IoU score of 0.956 and dice score of 0.972.

193

194         *Heatmaps:* To visualize the spatial location on the images that the CNN networks were

195   paying attention to, heatmaps were generated with class activation maps algorithm (38). This was

196   done by adding global average pooling into CNN and calculating gradient backpropagation given

197   one specific output class to obtain the class activation maps, indicating the discriminative image

198   regions CNN paid attention to.

199

200         *Statistical methods and performance evaluation:* Five-fold cross-validation was used for

201   the test set separately. Performance of the prediction model used standard ROC analysis of the

202   area under the curve (AUC), accuracy, sensitivity, specificity, precision, recall and F1 scores.

203   Precision was computed using true positives divided by the sum of false positives and true

204   positives; Recall was computed using the true positives divided by the sum of true positives and

205   false negatives; F1 scores were the mean of recall and precision rates.

206

## Results

207

208        **Figure 2** shows examples of pCXR from a normal subject, patients with bacterial

209 pneumonia, non-COVID-19 viral pneumonia, and COVID-19 infection. COVID-19 is often

210 characterized by ground-glass opacities with or without nodular consolidation with predominance

211 of bilateral, peripheral and lower lobes involvement. Non-COVID-19 viral pneumonia is often

212 characterized by diffuse interstitial opacities, usually bilaterally. Bacterial pneumonia is often

213 characterized by confluent areas of focal airspace consolidation.

214        **Figure 3** shows the training and validation loss and accuracy as a function of the epoch of

215 the CNN models. Loss decreases and accuracy improved with increasing epoch for both training

216 and validation dataset. The accuracy typically reached > 0.8.

217        CNN was used to classify COVID-19 pCXR from those of normal, bacterial pneumonia,

218 and non COVID-19 viral pneumonia patients in a multi-class neural network model. The results

219 of the multi-class CNN classification for the whole CXR in the form of the confusion matrix is

220 shown in **Table 1**. The precision, recall, and F1 scores for the whole pCXR (**Table 2**) showed

221 good to excellent performance. For CNN with transfer learning performed on the **whole** pCXR,

222 the overall sensitivity, specificity, accuracy, and AUC were 0.79, 0.93, and 0.79, .84 respectively.

223 For CNN performed on **segmented** lungs, the overall sensitivity, specificity, accuracy, and AUC

224 were 0.91, 0.93, 0.88, 0.89 respectively. The performance was generally better using segmented

225 lungs.

226        To visualize the spatial location on the images that the CNN networks were paying

227 attention to for classification, heatmaps of the COVID-19 versus normal pCXR are shown in

228 **Figure 4**. The CNN algorithm was able to localize the area of pathology on pCXR. For CNN

229 performed on the whole pCXR, the majority of the hot spots were reasonably localized to regions

230    of ground glass opacities and/or consolidations, but some hot spots were located outside the lungs.

231    For CNN performed on segmented lungs, the majority of the hot spots were reasonably localized

232    to regions of ground glass opacities and/or consolidations, mostly as expected. There were a few

233    pixels outside the lung that the algorithm paid attention to. These "errors" could be due to small sample

234    sizes. It learned from the training dataset and there may be information that the algorithm might consider

235    important. Large sample size usually minimizes such "error."

236

237    **Discussion**

238    This study developed and applied a deep-learning CNN algorithm with transfer learning to

239    classify COVID-19 CXR from normal, bacterial pneumonia, and non-COVID viral pneumonia

240    CXR in a multiclass model. Heatmaps showed reasonable localization of abnormalities in the

241    lungs. The overall sensitivity, specificity, accuracy, and AUC were 0.91, 0.93, 0.88, and .89

242    respectively (segmented lungs).

243

244    There are a few AI studies to date using machine learning methods to classify CXRs of

245    COVID-19, normal and related lung infections. By the time this paper is reviewed many more

246    papers will be published. **Hurt et al.** used a U-net CNN algorithm to predict pixel-wise probability

247    maps for pneumonia on CXR on 10 COVID-19 patients (Hurt et al. 2020). No ROC analysis was

248    performed. **Apostolopoulos and Mpesiana** used deep-learning algorithm to predict COVID-19

249    CXR with 98.66% sensitivity, 96.46% specificity, and 96.78% accuracy from a collection of 1427

250    CXRs of which 224 were COVID-19 CXRs (Apostolopoulos & Mpesiana 2020). **Elaziz et al.**

251    used an innovative feature selection algorithms and standard classifier to classify CXR between

252    COVID-19 (N=216) and non-COVID-19 (N=1675). This method achieved accuracy rates of

253    96.09% and 98.09% for each of the respective datasets (Elaziz et al. 2020). Note that patient

254    cohorts were highly asymmetric. **Murphy et al.** used an artificial intelligence to classify COVID-

255    19 CXRs (N=223) from non-COVID-19 CXRs (N=231) with an 0.81 AUC and they also showed

256    that AI outperformed expert readers (Murphy et al. 2020). **Ozturk et al.** used an AI model to

257    perform multiclass classification for COVID-19 (N=127) vs. No-Findings (N=500) vs. Pneumonia

258    (N=500) as well as a binary classification for COVID vs. No-Findings which achieved 87.02%

259    and 98.08% accuracies, respectively(Ozturk et al. 2020). **Pereira et al.** performed a multiclass

260    classification and a hierarchical classification for COVID-19 vs pneumonia vs no-finding using

261    resampling algorithms, texture descriptors, and CNN. This model achieved a F1-Score of 0.65 for

262    the multiclass approach and F1 score of 0.89 for the hierarchical classification (Pereira et al. 2020).

263    AUC and accuracy were not reported. AI has also been employed to stage pCXR disease severity

264    against radiologist scores (Cohen et al. 2020a; Zhu et al. 2020a).  Our study had one of the larger

265    cohorts, balanced sample sizes, and multi-class model. Our approach is also amongst the simplest

266    AI models with comparable performance index, likely facilitate immediate clinical translation.

267    Together, these studies indicate that AI has the potential to assist frontline physicians in

268    distinguishing COVID-19 infection based on CXRs.

269

270        Heatmaps are informative tools to visualize regions that CNN algorithm pays attention to

271    for detection. This is particular important given AI operates on high dimensional space. Such

272    heatmaps enable reality checks and make AI interpretable with respect to clinical findings. Our

273    algorithm showed that the majority of the hotspots were highly localized to abnormalities within

274    the lungs, i.e., ground glass opacity and/or consolidation, albeit imperfect. The majority of the

275    above-mentioned machine learning studies to classify COVID-19 CXRs did not provide heatmaps.

276    We also noted that CNN on whole pCXR image resulted in some hot spots located outside the

277   lungs. CNN of segmented lungs solved this problem. Another advantage of using segmented lung

278   is reduced computational cost during training. Transfer learning also reduced computational cost,

279   making this algorithm practical. The performance is generally better using segmented lungs.

280

281        Most COVID-19 positive patients showed significant abnormalities on pCXR. Some early

282   studies have even suggested that pCXR could be used as a primary tool for COVID-19 screening

283   in epidemic areas (Ai et al. 2020), which could complement swab testing which still has long

284   turnaround time and non-significant false positive rate. In some cases, imaging revealed chest

285   abnormalities even before swab tests confirm infection (Fang et al. 2020, in press; Li et al. 2020a).

286   In addition, pCXR can detect superimposed bacteria pneumonia, which necessitates urgent

287   antibiotic treatment. pCXR can also suggest acute respiratory distress syndrome, which is

288   associated with severe negative outcomes and necessitates immediate treatment. Together with the

289   potential widespread shortage of intensive care units and mechanical ventilators in many hospitals,

290   pCXR may play a critical role in decision-making. A timely implementation of AI methods could

291   help to realize the full potential of pCXR in this COVID-19 pandemic.

292

293        This pilot proof-of-principal study has several limitations. This is a retrospective study with

294   a small sample size and the data sets used for training had limited alternative diagnoses. Although

295   the Kaggle database has a large sample size for non-COVID-19 CXR, we chose the sample sizes

296   to be comparable to that of COVID-19 to avoid asymmetric sample sizes that could skew

297   sensitivity and specificity. Future studies will need to increase the COVID-19 sample size and

298   include additional lung pathologies. The spatiotemporal characteristics on pCXR of COVID-19

299   infection and its relation to clinical outcomes are unknown. Future endeavors could include

300  developing AI algorithms to stage severity, and predict progression, treatment response,

301  recurrence, and survival, to inform and advise risk management and resource allocation associated

302  with the COVID-19 pandemic, with inclusion of clinical variables in predictive models (Lam et

303  al. 2020; Zhao et al. 2020; Zhu et al. 2020b).

304

305      In conclusion, deep learning convolutional neural networks with transfer learning

306  accurately classify COVID-19 pCXR from pCXR of normal, bacterial pneumonia, and non-

307  COVID viral pneumonia patients in a multi-class neural nwork model. This approach has the

308  potential to help radiologists and frontline physicians by providing efficient and accurate

309  diagnosis.

310

311

312

313 **REFERENCES**

314

315 https://coronavirus.jhu.edu/map.html. *Johns Hopkin University*.

316 Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, Tao Q, Sun Z, and Xia L. 2020. Correlation of
317    Chest CT and RT-PCR Testing in Coronavirus Disease 2019 (COVID-19) in China: A
318    Report of 1014 Cases. *Radiology* 296:E32-E40.

319 Apostolopoulos ID, and Mpesiana TA. 2020. Covid-19: automatic detection from X-ray images
320    utilizing transfer learning with convolutional neural networks. *Phys Eng Sci Med* 43:635-
321    640. 10.1007/s13246-020-00865-4

322 Bao C, Liu X, Zhang H, Li Y, and Liu J. 2020. Coronavirus Disease 2019 (COVID-19) CT
323    Findings: A Systematic Review and Meta-analysis. *J Am Coll Radiol* 17:701-709.
324    10.1016/j.jacr.2020.03.006

325 Byra M, Styczynski G, Szmigielski C, Kalinowski P, Michalowski L, Paluszkiewicz R,
326    Ziarkiewicz-Wroblewska B, Zieniewicz K, Sobieraj P, and Nowicki A. 2018. Transfer
327    learning with deep convolutional neural network for liver steatosis assessment in
328    ultrasound images. *Int J Comput Assist Radiol Surg* 13:1895-1903. 10.1007/s11548-018-
329    1843-2

330 CDC. https://www.cdc.gov/coronavirus/2019-ncov/lab/index.html.

331 Cohen JP, Dao L, Roth K, Morrison P, Bengio Y, Abbasi AF, Shen B, Mahsa HK, Ghassemi M,
332    Li H, and Duong TQ. 2020a. Predicting COVID-19 Pneumonia Severity on Chest X-ray
333    With Deep Learning. *Cureus* 12:e9448. 10.7759/cureus.9448

334 Cohen JP, Morrison P, and Dao L. 2020b. COVID-19 Image Data Collection.
335    *arXiv:200311597v1*.

336 Ehteshami Bejnordi B, Mullooly M, Pfeiffer RM, Fan S, Vacek PM, Weaver DL, Herschorn S,
337    Brinton LA, van Ginneken B, Karssemeijer N, Beck AH, Gierach GL, van der Laak J,
338    and Sherman ME. 2018. Using deep convolutional neural networks to identify and
339    classify tumor-associated stroma in diagnostic breast biopsies. *Mod Pathol* 31:1502-
340    1512. 10.1038/s41379-018-0073-z

341 Elaziz MA, Hosny KM, Salah A, Darwish MM, Lu S, and Sahlol AT. 2020. New machine
342    learning method for image-based diagnosis of COVID-19. *PLoS One* 15:e0235187.
343    10.1371/journal.pone.0235187

344 Fang Y, Zhang H, Xie J, Lin M, Ying L, Pang P, and Ji W. 2020, in press. Sensitivity of Chest
345    CT for COVID-19: Comparison to RT-PCR. *Radiology*.

346 Harris M, Qi A, Jeagal L, Torabi N, Menzies D, Korobitsyn A, Pai M, Nathavitharana RR, and
347    Ahmad Khan F. 2019. A systematic review of the diagnostic accuracy of artificial
348    intelligence-based computer programs to analyze chest x-rays for pulmonary
349    tuberculosis. *PLoS One* 14:e0221339. 10.1371/journal.pone.0221339

350 Heo SJ, Kim Y, Yun S, Lim SS, Kim J, Nam CM, Park EC, Jung I, and Yoon JH. 2019. Deep
351    Learning Algorithms with Demographic Information Help to Detect Tuberculosis in
352    Chest Radiographs in Annual Workers' Health Examination Data. *Int J Environ Res*
353    *Public Health* 16. 10.3390/ijerph16020250

354 Holshue ML, DeBolt C, Lindquist S, Lofy KH, Wiesman J, Bruce H, Spitters C, Ericson K,
355    Wilkerson S, Tural A, Diaz G, Cohn A, Fox L, Patel A, Gerber SI, Kim L, Tong S, Lu X,
356    Lindstrom S, Pallansch MA, Weldon WC, Biggs HM, Uyeki TM, Pillai SK, and

357        Washington State -nCo VCIT. 2020. First Case of 2019 Novel Coronavirus in the United
358             States. *N Engl J Med* 382:929-936. 10.1056/NEJMoa2001191
359   Hui DS, E IA, Madani TA, Ntoumi F, Kock R, Dar O, Ippolito G, McHugh TD, Memish ZA,
360             Drosten C, Zumla A, and Petersen E. 2020. The continuing 2019-nCoV epidemic threat
361             of novel coronaviruses to global health - The latest 2019 novel coronavirus outbreak in
362             Wuhan, China. *Int J Infect Dis* 91:264-266. 10.1016/j.ijid.2020.01.009
363   Hurt B, Kligerman S, and Hsiao A. 2020. Deep Learning Localization of Pneumonia: 2019
364             Coronavirus (COVID-19) Outbreak. *J Thorac Imaging* 35:W87-W89.
365             10.1097/RTI.0000000000000512
366   Krizhevsky A, Sutskever I, and Hinton GE. 2012. Imagenet classification with deep
367             convolutional neural networks. Advances in neural information processing systems. p
368             1097-1105.
369   Lam KW, Chow KW, Vo J, Hou W, Li H, Richman PS, Mallipattu SK, Skopicki HA, Singer AJ,
370             and Duong TQ. 2020. Continued in-hospital ACE inhibitor and ARB Use in hypertensive
371             COVID-19 patients is associated with positive clinical outcomes. *J Infect Dis* 222:1256-
372             1264. 10.1093/infdis/jiaa447
373   LeCun Y, Bengio Y, and Hinton G. 2015. Deep learning. *Nature* 521:436-444.
374             10.1038/nature14539
375   Li D, Wang D, Dong J, Wang N, Huang H, Xu H, and Xia C. 2020a. False-Negative Results of
376             Real-Time Reverse-Transcriptase Polymerase Chain Reaction for Severe Acute
377             Respiratory Syndrome Coronavirus 2: Role of Deep-Learning-Based CT Diagnosis and
378             Insights from Two Cases. *Korean J Radiol* 21:505-508. 10.3348/kjr.2020.0146
379   Li K, Wu J, Wu F, Guo D, Chen L, Fang Z, and Li C. 2020b. The Clinical and Chest CT
380             Features Associated With Severe and Critical COVID-19 Pneumonia. *Invest Radiol*
381             55:327-331. 10.1097/RLI.0000000000000672
382   Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Ren R, Leung KSM, Lau EHY, Wong JY, Xing
383             X, Xiang N, Wu Y, Li C, Chen Q, Li D, Liu T, Zhao J, Liu M, Tu W, Chen C, Jin L,
384             Yang R, Wang Q, Zhou S, Wang R, Liu H, Luo Y, Liu Y, Shao G, Li H, Tao Z, Yang Y,
385             Deng Z, Liu B, Ma Z, Zhang Y, Shi G, Lam TTY, Wu JT, Gao GF, Cowling BJ, Yang B,
386             Leung GM, and Feng Z. 2020c. Early Transmission Dynamics in Wuhan, China, of
387             Novel Coronavirus-Infected Pneumonia. *N Engl J Med* 382:1199-1207.
388             10.1056/NEJMoa2001316
389   Lu H, Stratton CW, and Tang YW. 2020. Outbreak of pneumonia of unknown etiology in
390             Wuhan, China: The mystery and the miracle. *J Med Virol* 92:401-402.
391             10.1002/jmv.25678
392   Mekov E, Miravitlles M, and Petkov R. 2020. Artificial intelligence and machine learning in
393             respiratory medicine. *Expert Rev Respir Med* 14:559-564.
394             10.1080/17476348.2020.1743181
395   Murphy K, Smits H, Knoops AJG, Korst M, Samson T, Scholten ET, Schalekamp S, Schaefer-
396             Prokop CM, Philipsen R, Meijers A, Melendez J, van Ginneken B, and Rutten M. 2020.
397             COVID-19 on the Chest Radiograph: A Multi-Reader Evaluation of an AI System.
398             *Radiology*:201874. 10.1148/radiol.2020201874
399   Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, and Rajendra Acharya U. 2020.
400             Automated detection of COVID-19 cases using deep neural networks with X-ray images.
401             *Comput Biol Med* 121:103792. 10.1016/j.compbiomed.2020.103792

Pereira RM, Bertolini D, Teixeira LO, Silla CN, Jr., and Costa YMG. 2020. COVID-19 identification in chest X-ray images on flat and hierarchical classification scenarios. *Comput Methods Programs Biomed* 194:105532. 10.1016/j.cmpb.2020.105532

Ren T, Cattell R, Duanmu H, Huang P, Li H, Vanguri R, Liu MZ, Jambawalikar S, Ha R, Wang F, Cohen J, Bernstein C, Bangiyev L, and Duong TQ. 2020. Convolutional Neural Network Detection of Axillary Lymph Node Metastasis Using Standard Clinical Breast MRI. *Clin Breast Cancer* 20:e301-e308. 10.1016/j.clbc.2019.11.009

Samala RK, Chan HP, Hadjiiski LM, Helvie MA, Cha KH, and Richter CD. 2017. Multi-task transfer learning deep convolutional neural network: application to computer-aided diagnosis of breast cancer on mammograms. *Phys Med Biol* 62:8894-8908. 10.1088/1361-6560/aa93d4

Simonyan K, and A. Z. 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv:14091556*.

Song J, Hu B, Qu H, Wang L, Huang X, Li M, and Zhang M. 2020. Upregulation of angiotensin converting enzyme 2 by shear stress reduced inflammation and proliferation in vascular endothelial cells. *Biochem Biophys Res Commun* 525:812-818. 10.1016/j.bbrc.2020.02.151

Tang YW, Schmitz JE, Persing DH, and Stratton CW. 2020. Laboratory Diagnosis of COVID-19: Current Issues and Challenges. *J Clin Microbiol* 58. 10.1128/JCM.00512-20

Wang W, Xu Y, Gao R, Lu R, Han K, Wu G, and Tan W. 2020. Detection of SARS-CoV-2 in Different Types of Clinical Specimens. *JAMA* 323:1843-1844. 10.1001/jama.2020.3786

Xu X, Yu C, Qu J, Zhang L, Jiang S, Huang D, Chen B, Zhang Z, Guan W, Ling Z, Jiang R, Hu T, Ding Y, Lin L, Gan Q, Luo L, Tang X, and Liu J. 2020. Imaging and clinical features of patients with 2019 novel coronavirus SARS-CoV-2. *Eur J Nucl Med Mol Imaging* 47:1275-1280. 10.1007/s00259-020-04735-9

Zhao Z, Chen A, Hou W, Graham JM, Li H, Richman PS, Thode HC, Singer AJ, and Duong TQ. 2020. Prediction model and risk scores of ICU admission and mortality in COVID-19. *PLoS One* 15:e0236618. 10.1371/journal.pone.0236618

Zhu J, Shen B, Abbasi A, Hoshmand-Kochi M, Li H, and Duong TQ. 2020a. Deep transfer learning artificial intelligence accurately stages COVID-19 lung disease severity on portable chest radiographs. *PLoS One* 15:e0236621. 10.1371/journal.pone.0236621

Zhu JS, Ge P, Jiang C, Zhang Y, Li X, Zhao Z, Zhang L, and Duong TQ. 2020b. Deep-learning artificial intelligence analysis of clinical variables predicts mortality in COVID-19 patients. *J Am Coll Emerg Physicians Open*. 10.1002/emp2.12205

438

439    **Figure legends**

440

441    Figure 1. VGG16 architecture with 16 weighted layers including 3 fully connected layers.

442

443    Figure 2:  Examples of chest radiographs (a) normal, (b) COVID-19 viral pneumonia, (c) non-

444    COVID-19 viral pneumonia, and (d) bacterial pneumonia. COVID-19 is often characterized by

445    ground-glass opacities with or without nodular consolidation with predominance of bilateral,

446    peripheral and lower lobes involvement. Non-COVID-19 viral pneumonia is often characterized

447    by diffuse interstitial opacities, usually bilaterally. Bacterial pneumonia is often characterized by

448    confluent areas of focal airspace consolidation. Arrows indicate regions of above-described

449    characteristic features.

450

451    Figure 3:  CNN (a) training and (b) validation loss and accuracy. Loss decreases and accuracy

452    improved with increasing epoch for both training and validation dataset.

453

454    Figure 4:  (a) pCXR from a COVID-19 patient, (b) the corresponding segmented lung, (c) heatmap

455    from CNN analysis using whole pCXR, and (d) heatmap from CNN analysis using segmented lung

456    overlaid on whole CXR. Arrows indicated regions of ground glass opacity and/or consolidations.

457

Confusion table

**Table 1.** Confusion table showing the multiclass CNN classification (whole CXR)

1 **Table 1.** Confusion table showing the multiclass CNN classification (whole CXR)

|  | Normal | COVID-19 | Non-COVID-19 viral pneumonia | Bacterial pneumonia |
|---|---|---|---|---|
| **Normal** | 122 | 3 | 17 | 2 |
| **Covid19** | 6 | 102 | 3 | 6 |
| **Non-COVID-19 viral Pneumonia** | 16 | 2 | 94 | 20 |
| **Bacterial pneumonia** | 4 | 1 | 30 | 85 |

Precision and recall rate and F1 score (whole CXR).

**Table 2** shows the precision and recall rate and F1 score (whole CXR).

1        **Table 2** shows the precision and recall rate and F1 score (whole CXR).

|  | Precision | Recall | F1 - score |
|---|---|---|---|
| **Normal** | 0.82 | 0.85 | 0.84 |
| **Covid19** | 0.94 | 0.87 | 0.91 |
| **Non-covid19 viral pneumonia** | 0.65 | 0.71 | 0.68 |
| **Bacterial pneumonia** | 0.75 | 0.71 | 0.73 |

# Figure 1

VGG16 architecture

Figure 1. VGG16 architecture with 16 weighted layers including 3 fully connected layers.

**Input**

| Conv2D 1-1 |
| Conv2D 1-2 |
| Pooling |

| Conv2D 2-1 |
| Conv2D 2-2 |
| Pooling |

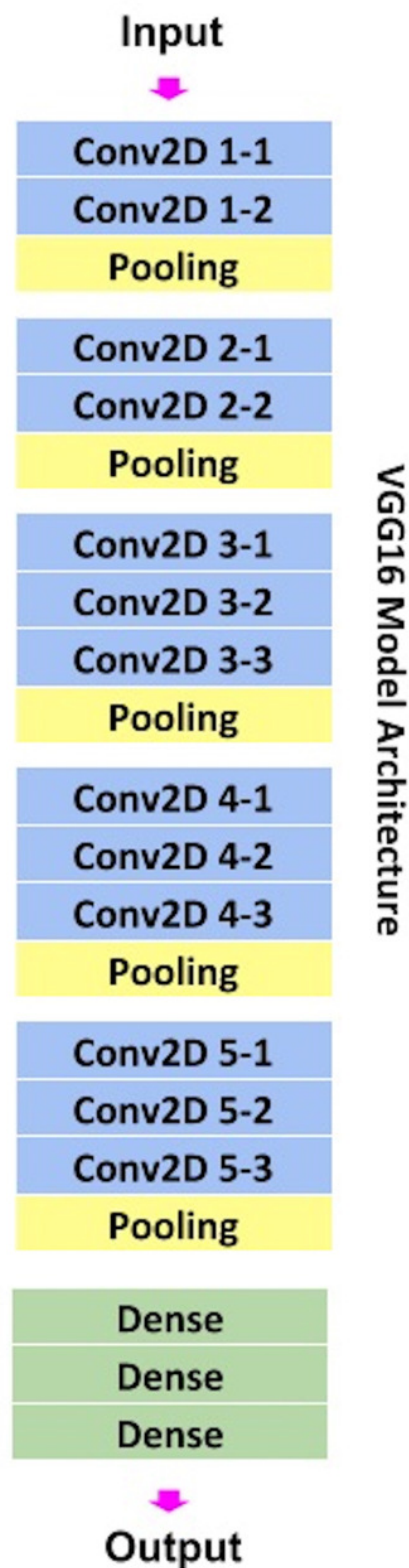
| Conv2D 3-1 |
| Conv2D 3-2 |
| Conv2D 3-3 |
| Pooling |

| Conv2D 4-1 |
| Conv2D 4-2 |
| Conv2D 4-3 |
| Pooling |

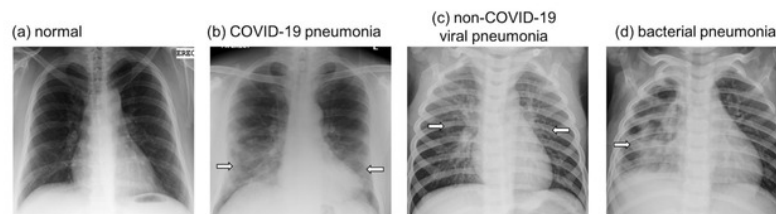
| Conv2D 5-1 |
| Conv2D 5-2 |
| Conv2D 5-3 |
| Pooling |

| Dense |
| Dense |
| Dense |

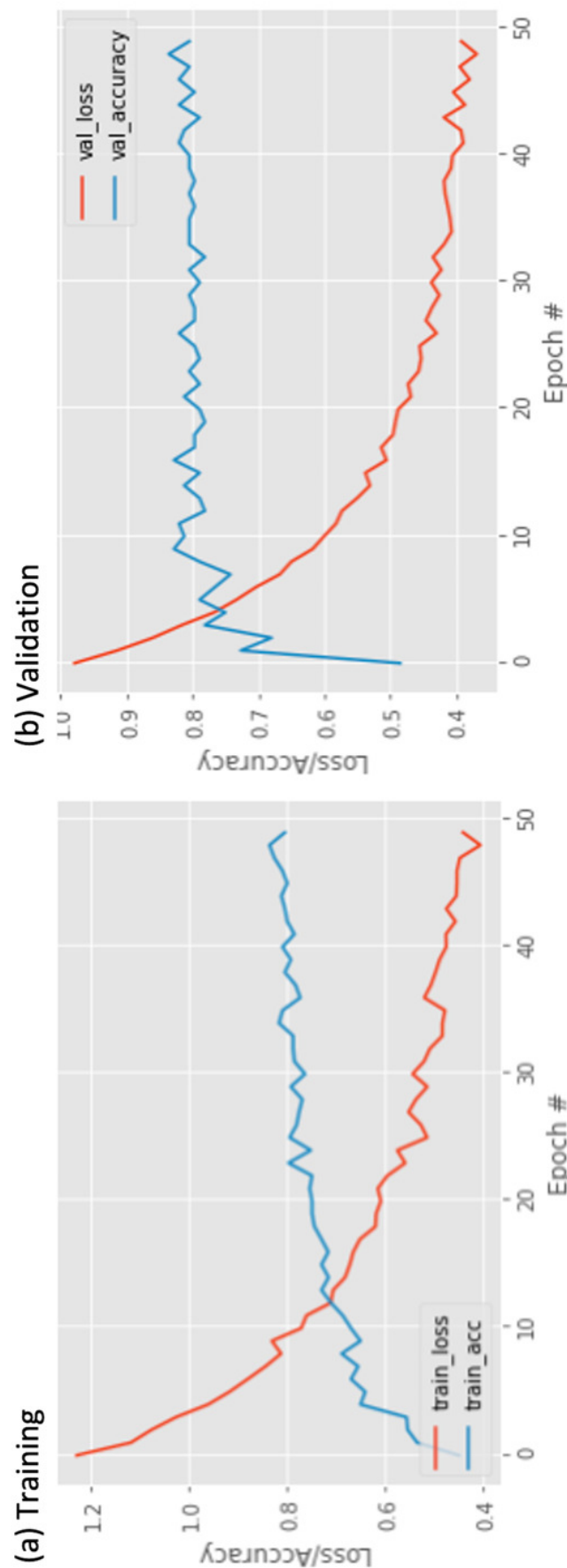
VGG16 Model Architecture

**Output**

# Figure 2

CXR

Figure 2: Examples of chest radiographs (a) normal, (b) COVID-19 viral pneumonia, (c) non-COVID-19 viral pneumonia, and (d) bacterial pneumonia. COVID-19 is often characterized by ground-glass opacities with or without nodular consolidation with predominance of bilateral, peripheral and lower lobes involvement. Non-COVID-19 viral pneumonia is often characterized by diffuse interstitial opacities, usually bilaterally. Bacterial pneumonia is often characterized by confluent areas of focal airspace consolidation. Arrows indicate regions of above-described characteristic features.

# Figure 3

CNN training and validation

Figure 3: CNN training and validation loss and accuracy. Loss decreases and accuracy improved with increasing epoch for both training and validation dataset.

# Figure 4

Heatmap

Figure 4: pCXR from a COVID-19 patient, the corresponding segmented lung, heatmap from CNN analysis using whole pCXR, and heatmap from CNN analysis using segmented lung overlaid on whole CXR. Arrows indicated regions of ground glass opacity and/or consolidations.

(d) Heatmap on segmented lung   (c) Heatmap on whole CXR   (b) Segmented lung   (a) CXR