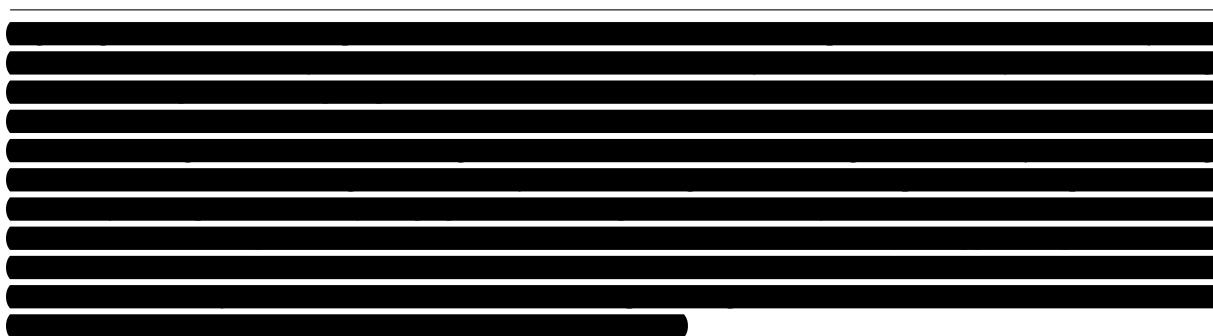


# A Deep Learning Approach for COVID-19 & Viral Pneumonia Screening with X-ray Images

FAIZAN AHMED, SYED AHMAD CHAN BUKHARI, and FAZEL KESHTKAR,  
St. John's University



CCS Concepts: • Computing methodologies → *Activity recognition and understanding*; Active learning settings;

Additional Key Words and Phrases: Deep learning, convolutional neural networks, computer vision, medical imaging, COVID-19

## ACM Reference format:

Faizan Ahmed, Syed Ahmad Chan Bukhari, and Fazel Keshtkar. 2020. A Deep Learning Approach for COVID-19 & Viral Pneumonia Screening with X-ray Images. *Digit. Gov.: Res. Pract.* 2, 2, Article 18 (December 2020), 12 pages.  
<https://doi.org/10.1145/3431804>

---

## 1 INTRODUCTION

COVID-19, the illness caused by the severe acute respiratory syndrome coronavirus (SARS-CoV-2) enters your body through the mucous membranes of your face, such as your nose, mouth, and eyes by latching its surface proteins to receptors on healthy cells [1]. Gradually, the virus spreads down the respiratory tract, which can cause patients' lungs to become inflamed and lead to difficulty in breathing. This can eventually develop into pneumonia, an infection of the *alveoli* inside the lung where the blood exchanges oxygen and carbon dioxide. Pneumonia occurs when parts of the lungs consolidate and collapse, as a result of an accumulation of pus in the

---

Authors' addresses: F. Ahmed, St. John's University, 8000 Utopia Pkwy, Jamaica, NY, 11439; email: faizan.ahmed18@stjohns.edu; S. A. C. Bukhari and F. Keshtkar, St. John's University; emails: {bukharis, keshtkarf}@stjohns.edu.



[This work is licensed under a Creative Commons Attribution International 4.0 License.](#)

© 2020 Copyright held by the owner/author(s).  
 2639-0175/2020/12-ART18  
<https://doi.org/10.1145/3431804>



Fig. 1. Comparison between each category of X-ray.

infected airways. Patients also suffer from fever and cough due to the reduced surfactant in the alveoli from the viral attack, which diminishes oxygen uptake [2, 3].

If a doctor performs a chest CT scan of a patient with COVID-19 or pneumonia, they will likely see shadows or patchy areas called “ground-glass opacity” and are extremely helpful in the early detection and diagnosis of this disease [4]. However, differentiating between the occurrence of the two diseases can be inherently difficult, which is what inspired a deep learning approach to this problem. Currently, the most common diagnostic test used to identify positive cases of COVID-19 is the reverse-transcription polymerase chain reaction test, or RT-PCR. The procedure works by detecting viral RNA in a person’s cells, usually collected from their nose. These tests can be highly inaccurate, as scientists from Johns Hopkins Medicine have shown that as many as 20% of RT-PCR tests may produce false negatives, incorrectly indicating that a patient does not have COVID-19, when they actually do [5]. In fact, RT-PCR tests generally do not perform well outside of ideal conditions, as researchers at the Foundation for Innovative New Diagnostics, a nonprofit research center in Geneva, Switzerland, achieved 100% sensitivity and at least 96% specificity on negative samples in a laboratory setting. However, with the abnormal and irregular testing conditions in the real world, clinical sensitivity of RT-PCR tests ranges from 66% to 80% [6]. On the basis of these findings, scientists state that it is important to exercise caution when interpreting results of RT-PCR tests for SARS-CoV-2, especially if the test took place in the early stages of infection.

Given the widespread presence of radiology imaging systems in modern healthcare systems, employing a deep learning approach to this problem can help increase efficiency in detecting positive cases of the disease. A trained deep learning model can be quickly deployed to detect abnormalities in the X-ray images and help differentiate between occurrences of COVID-19 and Viral Pneumonia. In fact, chest X-rays are often performed as part of a standard procedure for patients with a respiratory complaint, and they can be a great complement to the traditional RT-PCR testing to reach a diagnosis [7]. In the event that chest X-rays are used to screen for a diagnosis, the presence of expert radiologists would be required to interpret the images. However, given that visual indicators of COVID-19 and Viral Pneumonia can be quite elusive (as shown in Figure 1), even to experts, there is still a high chance of a misdiagnosis. Deep learning provides a fast, automated, and effective strategy to eliminate this problem, with an accuracy potentially higher than human capability. The computer learns to find the minuscule features in the X-rays that contribute to the diagnosis after analyzing hundreds of images, a process that can be very time-consuming for a radiologist to manually perform. The proposed model can be found at <https://github.com/faizancodes/COVID-19-X-Ray-Classification> and can potentially be utilized to screen for COVID-19 and assist physicians in their diagnosis of the patient.

## 2 DATASET AND IMAGES

In this research, the X-ray images were acquired from two Kaggle datasets [8, 9] that were created in collaboration with medical doctors to establish an open-source database to help researchers better understand COVID-19. The

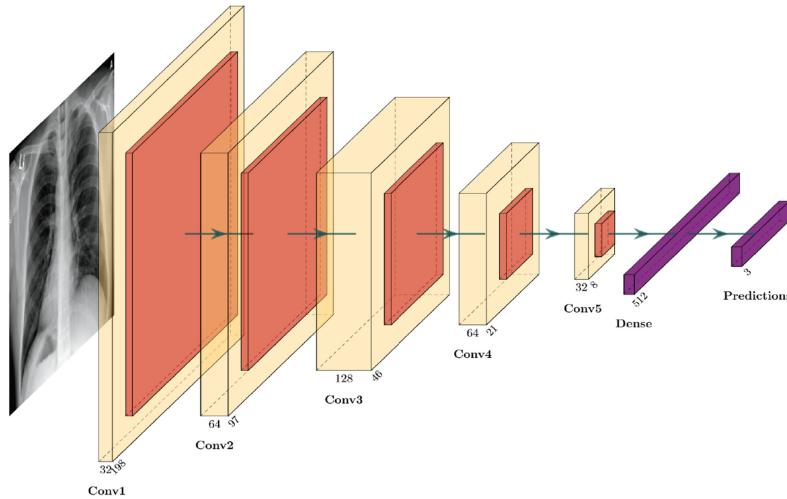


Fig. 2. Model architecture.

databases were created by extracting the X-ray images from various public sources such as publications as well as indirect collection from hospitals and physicians [10]. The dataset used in this article contains a total of 1,389 images, which includes 289 COVID-19, 550 Viral Pneumonia, and 550 Normal X-rays.

### 3 METHODS

The proposed model contains five convolutional layers, with each being followed by batch normalization and max pooling layers, along with dropout. Processing the final convolutional layer is a fully connected layer with 512 neurons, followed by the last layer with three neurons representing each category of X-ray. ReLU was used as the activation function for each layer, and softmax was used for the final dense layer. The model architecture is shown in Figure 2 and details for each layer are shown in Table 1.

The number of filters in each convolutional layer was gradually increased from 32 to 64 to 128, then back down to 64 and 32, with the strides for each layer being  $3 \times 3$ . Batch normalization was used to standardize the inputs, and max pooling was used to provide spacial variance, used to account for varying appearances.

The model was initially trained on a dataset containing 200 COVID-19 X-rays, 250 Viral Pneumonia X-rays, and 250 Normal X-rays, with each image being resized to a height and width of 200 by 200 pixels. Eighty percent of the data were used for training and 20% were used for validation, with 560 images being part of the training set and 140 images being part of the validation set. The developed model consists of 453,027 parameters and uses categorical crossentropy for the loss function, rmsprop for the optimizer, and a batch size of 5. The learning rate was initially set to 0.0001, and to prevent the accuracy from plateauing during training, we performed a learning rate reduction by a factor of 0.5 if the validation accuracy does not increase for two steps.

For the initial procedure, the model was tested on a dataset containing 89 COVID-19 X-rays, 250 Viral Pneumonia, and 250 normal X-rays. To further evaluate the model, we performed 5-fold cross-validation on the entire dataset of 1,389 X-ray images, where 80% of the dataset was used for training and 20% percent was used for testing. Of the training set, 20% was used as the validation set, with 889 total images used for training, 223 images used for validation, and 277 images used for testing for each fold of cross-validation. A visual representation of this procedure is shown in Figure 3 and Figure 4.

Table 1. Description of Each Layer in the Model

| Layer Type          | Output Shape         | Parameters |
|---------------------|----------------------|------------|
| Conv2D              | (None, 198, 198, 32) | 896        |
| Batch Normalization | (None, 198, 198, 32) | 128        |
| MaxPooling2D        | (None, 99, 99, 32)   | 0          |
| Dropout             | (None, 99, 99, 32)   | 0          |
| Conv2D              | (None, 97, 97, 64)   | 18,496     |
| Batch Normalization | (None, 97, 97, 64)   | 256        |
| MaxPooling2D        | (None, 48, 48, 64)   | 0          |
| Dropout             | (None, 48, 48, 64)   | 0          |
| Conv2D              | (None, 46, 46, 128)  | 73,856     |
| Batch Normalization | (None, 46, 46, 128)  | 512        |
| MaxPooling2D        | (None, 23, 23, 128)  | 0          |
| Dropout             | (None, 23, 23, 128)  | 0          |
| Conv2D              | (None, 21, 21, 64)   | 73,792     |
| Batch Normalization | (None, 21, 21, 64)   | 256        |
| MaxPooling2D        | (None, 10, 10, 64)   | 0          |
| Dropout             | (None, 10, 10, 64)   | 0          |
| Conv2D              | (None, 8, 8, 32)     | 18,464     |
| Batch Normalization | (None, 8, 8, 32)     | 128        |
| Max Pooling2D       | (None, 4, 4, 32)     | 0          |
| Dropout             | (None, 4, 4, 32)     | 0          |
| Flatten             | (None, 512)          | 0          |
| Dense               | (None, 512)          | 262,656    |
| Batch Normalization | (None, 512)          | 2,048      |
| Dropout             | (None, 512)          | 0          |
| Predictions (Dense) | (None, 3)            | 1,539      |

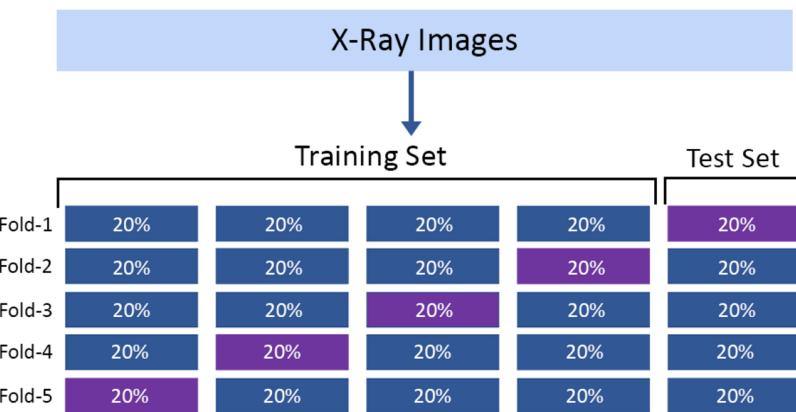


Fig. 3. Schematic representation of the 5-fold cross-validation procedure.

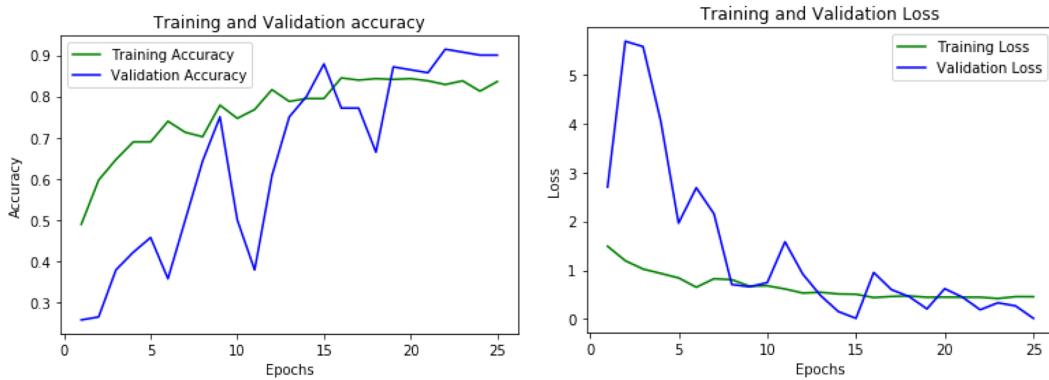


Fig. 4. Training, validation accuracy, and loss of the model after 25 epochs.

Table 2. Model Performance (%) for Each Fold of Cross-validation

| Fold    | Precision | Recall | F-1 Score | Accuracy |
|---------|-----------|--------|-----------|----------|
| Fold-1  | 91        | 91     | 91        | 91.37    |
| Fold-2  | 92        | 88     | 89        | 89.57    |
| Fold-3  | 94        | 89     | 91        | 91.73    |
| Fold-4  | 92        | 87     | 88        | 90.29    |
| Fold-5  | 91        | 90     | 90        | 90.25    |
| Average | 92        | 89     | 89.8      | 90.64    |

To better account for variations in the X-ray images, the `ImageDataGenerator` class in Keras<sup>1</sup> was used to generate batches of image data with real-time augmentation. The rotation, shear, zoom, width shift, and height shift ranges were all adjusted to account for variations in the images and best train the model.

## 4 RESULTS

For the initial procedure, the proposed model achieved an accuracy of 93% on the testing set, with an F1-score of 93%. The confusion matrix of the model performance is shown in Figure 5, which indicates that the model is successfully able to differentiate between the occurrence of the diseases, with high precision and recall for each category of X-ray.

The model performed similarly for each fold of cross-validation, with an average accuracy of 90.64% and F-1 Score of 89.88%, as shown in Table 2 and the confusion matrices in Figure 6.

## 5 DISCUSSION

Using saliency maps, we can better understand the features in the X-rays and visualize what areas of the image are of high importance. To visualize the saliency, the gradient of the output category was computed with respect to the input image. This tells us how the output category value changes with respect to a small change in input image's pixels. All the positive values in the gradients tell us that a small change to that pixel will increase the output value. Visualizing these gradients, which are the same shape as the image, should provide some insight on the most important features of the image [11].

<sup>1</sup><https://keras.io/api/preprocessing/image/>.

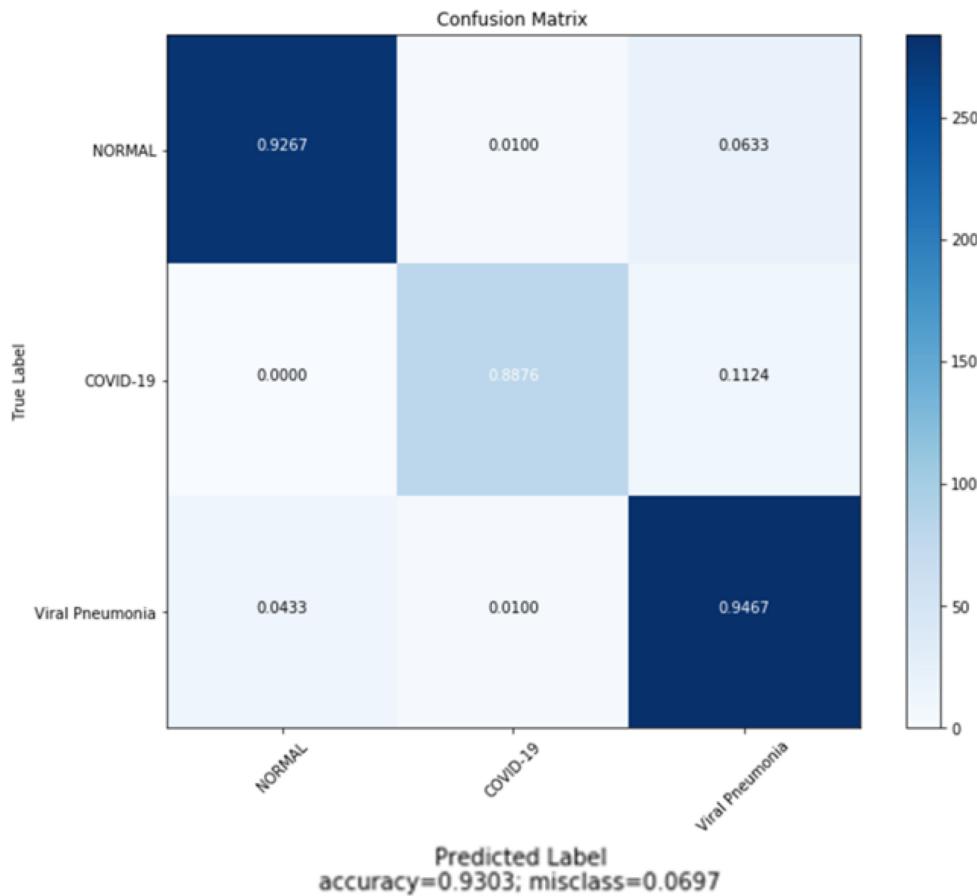


Fig. 5. Confusion matrix for model performance on testing set.

As shown in Figures 7–9, the areas of yellow gradient have the greatest influence on the model's prediction. These visualizations can be shown to radiologists to confirm their significance, as they help us better interpret the trained model.

From evaluating the COVID-19 saliency maps in Figure 7, it is evident that the areas around the cardiophrenic angle are what the model deems most important, along with areas around the left and right hilum. The areas along the azygo-esophageal recess and right middle thorax are also important in screening for all three categories of X-rays as well, as indicated by their respective saliency maps. These areas can be classified as containing patches of ground-glass opacity, which is critical for early detection of COVID-19 and Viral Pneumonia, hence the absence of the patches in the normal X-rays. Radiologists can analyze these visualizations to better understand each disease and discover new ways to screen for them.

Despite achieving a high accuracy with our deep learning model, there are several factors that prevent it from being adapted in the healthcare industry as of yet. The biggest issue we are faced with is the lack of sufficient data for COVID-19, since it is still a new disease. There are limited datasets available for X-ray scans of COVID-19 patients, and to determine the diagnostic performance of our model, a clinical study would have to be conducted. Due to the small dataset used, there can be hidden biases in the data, especially with no metadata on age, gender, different pathologies also present in the patients, and other information necessary to spot these kinds of biases.

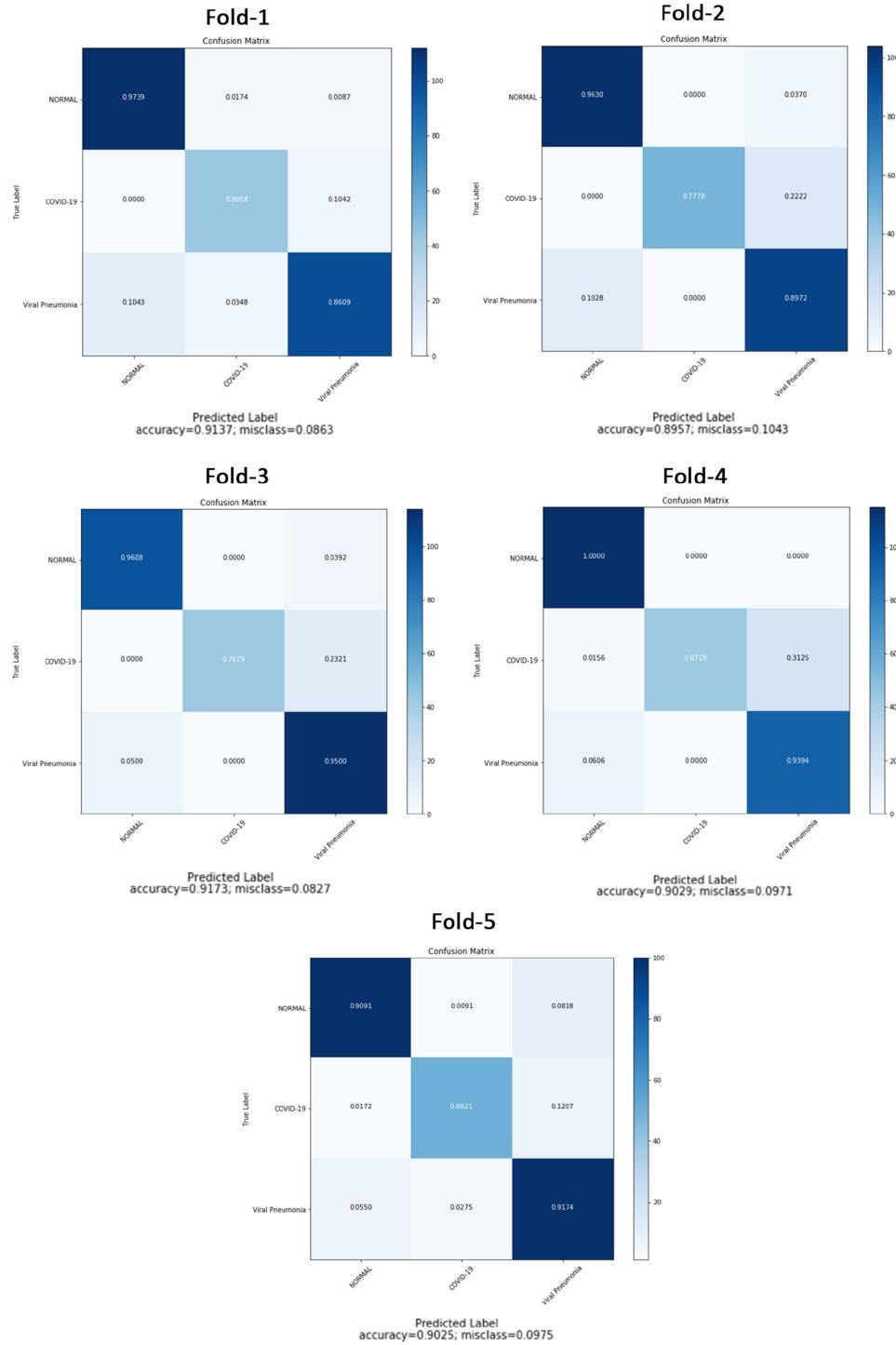


Fig. 6. Confusion matrices for each fold of the cross-validation procedure.

## COVID-19

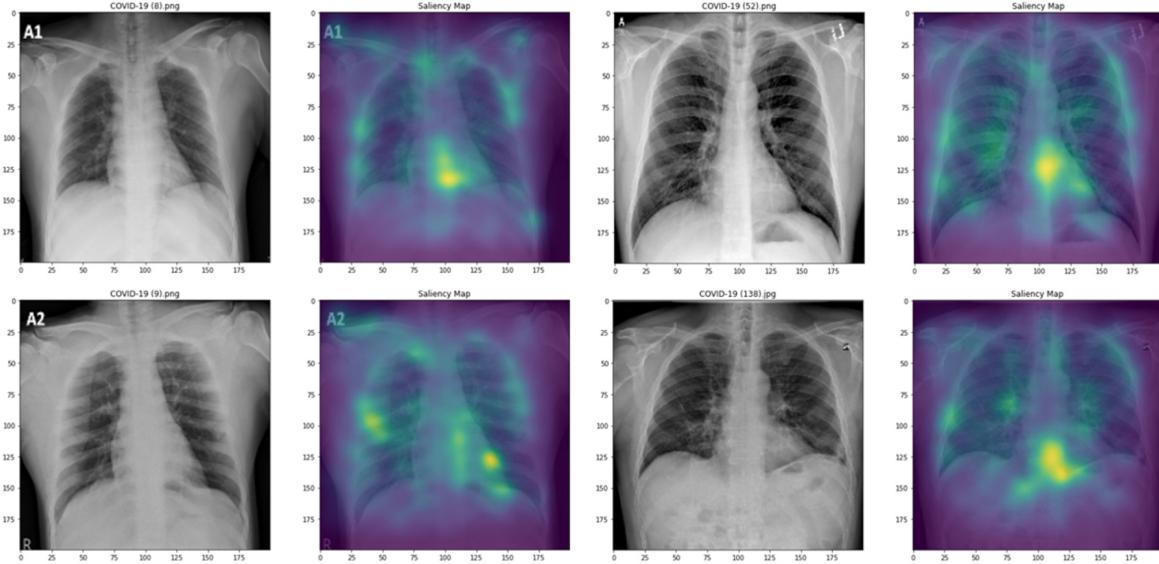


Fig. 7. Saliency maps highlighting areas of high importance for COVID-19 X-rays.

## Viral Pneumonia

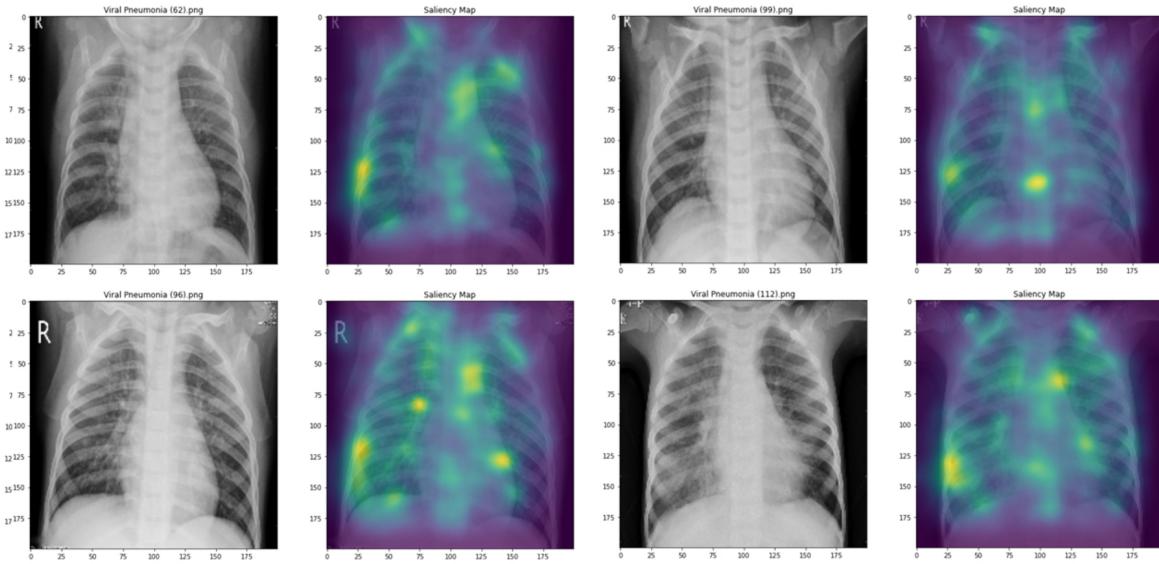


Fig. 8. Saliency maps highlighting areas of high importance for Viral Pneumonia X-rays.

## Normal

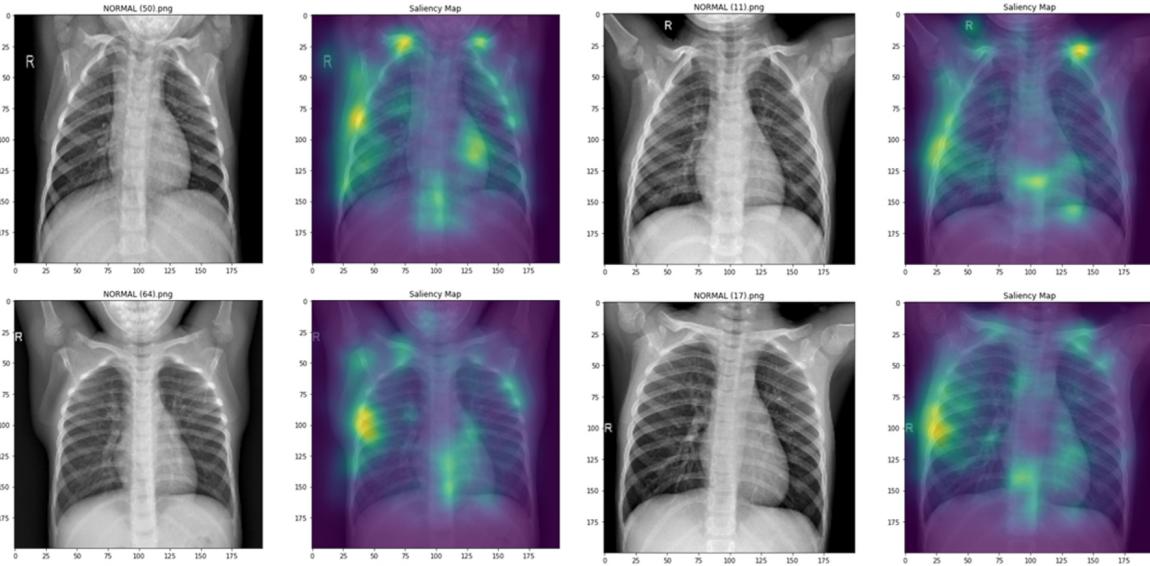


Fig. 9. Saliency maps highlighting areas of high importance for normal X-rays.

[12]. It was also found that convolutional neural networks are learning patterns in the X-ray images that are not correlated to the presence of COVID-19 at all, where similar classification performance can be obtained using X-ray images that do not contain most of the lungs [13]. Until COVID-19 data become more widely available and clinical studies are conducted, the proposed model cannot be used solely as a diagnostic measure.

## 6 RELATED WORKS

COVID-19 is a new phenomenon for scientists, especially in the AI community. However, some research has been done that focuses on image processing and chest X-ray images. In this section, we explain the current research that has used AI and machine learning techniques for X-rays and chest area cancer detection.

### 6.1 Pulmonary Disease Detection

In Jin et al. [14], their method achieved a detection rate of 92.4%, as their model was used to distinguish between healthy and pathological crackles. They used supervised machine learning where *K*-NN classification performed the best.

Flietstra et al. [15] used Support Vector Machine algorithms to detect heart failure between pneumonia and congestive. They used a dataset that contained 257 patients, and their method performance ranged from 82% to 87%.

Research done by Lang et al. [16] performed with f-measure of 0.90. They used semi-supervised and graph based approaches and applied SVM classification to distinguish between normal and abnormal lung sounds.

Torre-Cruz et al. [17] proposed a system to find the presence of wheeze sounds in breath recordings. Their model achieved 95.5% accuracy for classifying between presence and absence cases.

## 6.2 COVID-19 Disease Methods

Extensive research has been conducted that explores on COVID-19 detection. We provide a few of these research such as: References [18–20] (this is direction of the method we propose) and to protein analysis to boost research in COVID-19 vaccines, References [21] and [22].

Brunese et al. [18] introduced a method to reduce the time window to obtain a COVID-19 diagnosis to 2.5 seconds. Their method is based on transfer learning using the VGG-16 model. To do this, they built two models. The first model's goal was to detect whether a chest X-ray is related to a patient with generic pulmonary disease. If first case is true, then they consider this as input the X-ray to a second model. The second model aims to detect whether the pulmonary disease is COVID-19. Based on their results, where they used a total of 6,523 chest X-rays, their method obtained an accuracy of 0.96 for the discrimination between healthy and generic pulmonary disease patients and an accuracy of 98% for the COVID-19 detection.

XU et al. in References [19] and [36] applied deep learning using the UNet++ model and obtained 98.85% accuracy. They used 51 patients confirmed as affected by the COVID-19 disease from institutions such as Wuhan University. The authors in References [19] and [36] did not consider normal patients without COVID-19.

Xu et al. in Reference [20] improved their previous method [19] using deep learning and model performed an accuracy of 86.7%. They used 618 medical images for their model. They used two three-dimensional CNNs: ResNet23 network, the second network representing a variant of the first model, where they applied several layers.

In Reference [21], Beck et al. applied deep learning techniques in proteins to find new vaccines for COVID-19. Beck et al. [39] model used SMILES dataset, a repository of molecules as text to encode and decode each molecule. They found that the COVID-19 is 3C-like proteinase and binds with atazanavir.

Reference [22] investigated a COVIDX-Net model for COVID-19 to analyze X-ray images. Their model obtained 90% accuracy by using 25 COVID-19 positive and 25 non-COVID-19 images.

From above research, we consider that our method used various and different datasets of COVID-19 and non-COVID-19 chest X-ray images, which allowed our method to differentiate between the instances of the disease with high accuracy. Finally, we explore the saliency maps to help make the diagnostics explainable.

## 7 CONTRIBUTION AND IMPACT ON SOCIETY

We believe our system will be helpful to detect and determine X-ray images from patients that have similar characteristics to other diseases where symptoms imaging has similar patterns, i.e., COVID-19 and other illness categories. We also expect our system will detect different imaging features that infected with COVID-19 disease or similar disease such as flu ground. Our system especially will be useful to detect early stages of an outbreak and can be used in various government organizations or clinics that use X-ray images.

We also expect, if our system is used in a healthcare organization that is facing a new outbreak, healthcare professionals, doctors, physicians, and admin offices could be alerted due to high accuracy of our model. Our system will assist to provide more hints and signals on outbreak and identifying further image-based indication and classification of infectious images. Using this system also delivers more awareness into patient information system in organization for future usage by prognosis and treatment.

Finally, with the number of positive cases of the disease rising in cities worldwide and limited resources available in hospitals, our proposed deep learning model can potentially be used to screen for COVID-19 and detect instances of the disease much faster and more accurately than the traditional RT-PCR testing method, which can have high rates of false negative results. By detecting and isolating those who are infected with COVID-19 faster, cities can greatly benefit from lower rates of infection and be able to recover their economy as a whole.

## 8 CONCLUSION

After evaluating the model's performance, it is apparent that deep learning can successfully be used to screen for COVID-19 and Viral Pneumonia in chest X-rays. With an average classification accuracy of 90.64% and F-Score of 89.88% after performing 5-fold cross-validation, the model offers an automated, end-to-end solution for testing. It can potentially be used as an alternative or in conjunction with the standard RT-PCR testing methods currently used to increase efficiency and efficacy in detecting positive cases of each disease. In countries and areas where there are limited COVID-19 testing kits, utilizing this deep learning approach for testing can drastically help prevent the spread of the disease. However, due to the shortage of COVID-19 X-rays used, we aim to apply this model on a larger dataset to determine its true efficacy. With a clinical study, the true diagnostic performance of the model can be interpreted to then be used in healthcare facilities around the world.

## REFERENCES

- [1] Queensland Health. 2020. How does COVID-19 spread and how can I stop myself from getting it? Retrieved from <https://www.health.qld.gov.au/news-events/news/novel-coronavirus-covid-19-how-it-spreads-transmission-infection-prevention-protection>.
- [2] Hilary Guite. 2020. COVID-19: What happens inside the body? Retrieved from <https://www.medicalnewstoday.com/articles/covid-19-what-happens-inside-the-body>.
- [3] Jocelyn Kaiser Meredith Wadman, Jennifer Couzin-Frankel and Catherine Matacic. 2020. How does coronavirus kill? Clinicians trace a ferocious rampage through the body, from brain to toes. Retrieved from <https://www.sciencemag.org/news/2020/04/how-does-coronavirus-kill-clinicians-trace-ferocious-rampage-through-body-brain-toes>.
- [4] W. Yang, A. Sirajuddin, and X. Zhang. 2020. The role of imaging in 2019 novel coronavirus pneumonia (COVID-19). Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7156903/>.
- [5] Eleanor Bird. 2020. Tests may miss more than 1 in 5 COVID-19 cases. Retrieved from <https://www.medicalnewstoday.com/articles/tests-may-miss-more-than-1-in-5-covid-19-cases>.
- [6] Emily Waltz. April 2020. Testing the tests: Which COVID-19 tests are most accurate? Retrieved from <https://spectrum.ieee.org/the-human-os/biomedical/diagnostics/testing-tests-which-covid19-tests-are-most-accurate>.
- [7] A. Nair et al. 2020. A British Society of Thoracic Imaging statement: Considerations in designing local imaging diagnostic algorithms for the COVID-19 pandemic. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7156903/>.
- [8] Muhammad E. H. Chowdhury, Tawsifur Rahman, Amith Khandakar, Rashid Mazhar, Muhammad Abdul Kadir, Zaid Bin Mahbub, Khandaker Reajul Islam, Muhammad Salman Khan, Atif Iqbal, Nasser Al-Emadi, Mamun Bin Ibne Reaz, and T. I. Islam. 2020. Can AI help in screening Viral and COVID-19 pneumonia? Retrieved from <https://www.kaggle.com/tawsifurrahman/covid19-radiography-database>.
- [9] Nabeel Sajid. 2020. COVID-19 Patients Lungs X Ray Images 10000. Retrieved from <https://www.kaggle.com/nabeelsajid917/covid-19-x-ray-10000-images>.
- [10] Joseph Paul Cohen, Paul Morrison, and Lan Dao. 2020. COVID-19 image data collection. Retrieved from <https://github.com/ieee8023/covid-chestxray-dataset>.
- [11] Raghavendra Kotikalapudi and contributors. 2017. keras-vis. Retrieved from <https://github.com/raghakot/keras-vis>.
- [12] Enzo Tartaglione, Carlo Alberto Barbano, Claudio Berzovini, Marco Calandri, and Marco Grangetto. 2020. Unveiling COVID-19 from CHEST X-Ray with deep learning: A hurdles race with small data. *Int. J. Environ. Res. Pub. Health* 17, 18 (Sep. 2020), 6933. DOI: <http://dx.doi.org/10.3390/ijerph17186933>
- [13] Gianluca Maguolo and Loris Nanni. 2020. "A Critic Evaluation of Methods for COVID-19 Automatic Detection from X-Ray Images." arxiv:eess.IV/2004.12823.
- [14] F. Jin, S. Krishnan, and F. Sattar. 2011. Adventitious sounds identification and extraction using temporal-spectral dominance-based features. *IEEE Trans. Biomed. Eng.* 58, 11 (2011), 3078–3087.
- [15] B. Flietstra, N. Markuzon, A. Vyshedskiy, and R. Murphy. 2011. Automated analysis of crackles in patients with interstitial pulmonary fibrosis. *Pulmon. Med.* (2011). DOI: <http://dx.doi.org/10.1155/2011/590506>
- [16] Rongling Lang, Ruibo Lu, Chenqian Zhao, Honglei Qin, and Guodong Liu. 2020. Graph-based semi-supervised one class support vector machine for detecting abnormal lung sounds. *Appl. Math. Comput.* 364 (2020), 124487. DOI: <http://dx.doi.org/10.1016/j.amc.2019.06.001>
- [17] J. Torre-Cruz, F. Canadas-Quesada, S. García-Galán, N. Ruiz-Reyes, P. Vera-Candeas, and J. Carabias-Orti. 2020. A constrained tonal semi-supervised non-negative matrix factorization to classify presence/absence of wheezing in respiratory sounds. *Appl. Acoust.* 161 (2020), 107188. DOI: <http://dx.doi.org/10.1016/j.apacoust.2019.107188>
- [18] Luca Brunese, Francesco Mercaldo, Alfonso Reginelli, and Antonella Santone. 2020. Explainable deep learning for pulmonary disease and coronavirus COVID-19 detection from X-rays. *Comput. Meth. Prog. Biomed.* 196 (2020), 105608. DOI: <http://dx.doi.org/10.1016/j.cmpb.2020.105608>

- [19] Xiaowei Xu, Xiangao Jiang, Chunlian Ma, Peng Du, Xukun Li, et al. 2020. Deep learning-based model for detecting 2019 novel coronavirus pneumonia on high-resolution computed tomography: A prospective study. *Engineering* (2020). DOI :<http://dx.doi.org/10.1016/j.eng.2020.04.010>
- [20] Xiaowei Xu, Xiangao Jiang, Chunlian Ma, Peng Du, Xukun Li, et al. 2020. A deep learning system to screen novel coronavirus disease 2019 pneumonia. *Engineering* (2020). DOI :<http://dx.doi.org/10.1016/j.eng.2020.04.010>
- [21] Bo Ram Beck, Bonggun Shin, Yoonjung Choi, Sungsoo Park, and Keunsoo Kang. 2020. Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. *Computat. Struct. Biotechnol. J.* 18 (2020), 784–790. DOI :<http://dx.doi.org/10.1016/j.csbj.2020.03.025>
- [22] Ezz El-Din Hemdan, Marwa A. Shouman, and Mohamed Esmail Karar. 2020. COVIDX-Net: A Framework of Deep Learning Classifiers to Diagnose COVID-19 in X-Ray Images. arxiv:eess.IV/2003.11055.

Received July 2020; revised October 2020; accepted October 2020