An Explainable AI Model for Interpretable Lung Disease Classification

Vidhi Pitroda*1, Mostafa M. Fouda†2, and Zubair Md Fadlullah*‡3.
*Department of Computer Science, Lakehead University, Thunder Bay, Ontario, Canada.
†Department of Electrical and Computer Engineering, Idaho State University, Pocatello, ID, USA.
‡Thunder Bay Regional Health Research Institute (TBRHRI), Thunder Bay, Ontario, Canada.
Emails: 1pitrodavidhi7@gmail.com, 2mfouda@ieee.org, 3zubair.fadlullah@lakeheadu.ca.

***Abstract*—In this paper, we develop a framework for lung disease identification from chest X-ray images by differentiating the novel coronavirus disease (COVID-19) or other disease-induced lung opacity samples from normal cases. We perform image processing tasks, segmentation, and train a customized Convolutional Neural Network (CNN) that obtains reasonable performance in terms of classification accuracy. To address the black-box nature of this complex classification model, which emerged as a key barrier to applying such Artificial Intelligence (AI)-based methods for automating medical decisions raising skepticism among clinicians, we address the need to quantitatively interpret the performance of our adopted approach using a Layer-wise Relevance Propagation (LRP)-based method. We also used a pixel flipping-based, robust performance metric to evaluate the explainability of our adopted LRP method and compare its performance with other explainable methods, such as Local Interpretable Model Agnostic Explanation (LIME), Guided Backpropagation (GB), and Deep Taylor Decomposition (DTD).**

***Index Terms*—Deep learning, explainable AI, Layer-wise Relevance Propagation, LIME, Deep Taylor Decomposition, Guided Backpropagation, medical diagnosis, chest X-ray, COVID-19.**

I. Introduction

The power and benefit of Artificial Intelligence (AI) techniques, such as machine and deep learning models, have recently demonstrated a remarkable advancement in medical diagnosis; e.g., automated classification of heart disease [1], classification of diabetic retinopathy [2], tumor detection, nodule identification [3], and so forth. These techniques have significantly improved early diagnosis and have exhibited promising possibilities in reducing medical errors. Lungs plays a vital role in maintaining health of an individual. In a standard scenario, healthy lungs are full of air, and they look dark in the X-ray because there is no blockage in the lungs. In case of opacity, it starts filling partially with fluid or cells, and the space between the lungs starts thickening. It can sometimes result in short term illness or can also get serious. Thus, its early detection is essential to prevent it from getting severe. Further, it is also evident that the spread of COVID-19 has got a pernicious effect on people's health and well-being worldwide [4], [5].

Several deep learning-based approaches to detect lung opacity due to various lung diseases including the novel coronavirus disease (COVID-19), and the classification performances of these models have been remarkably high [6], [7]. This has, however, drawn both optimism and skepticism from the clinicians because of the shortcoming of these models to provide transparency and interpretability. In other words, while machine learning methods such as decision trees, random forest, and linear regression can be understood by decision boundary through visualization of the model parameters, it is not feasible in multi-dimensional classification in the aforementioned ”blackbox” models. The use of AI systems in industries are affected as it cannot provide explainability and reliability to the customer due to the absence of tools to examine the functioning of black-box models. Therefore, in this paper, we address the critical need for providing a human-like explainability and understanding of the model performance to promote why and how these models provide lung disease classification to encourage clinicians and radiologists embrace the automated AI classification systems.

In this vein, we first employ a chest X-ray dataset for lung opacity and COVID-19 detection. We perform image processing and segmentation to obtain the significant features for model training. We train a customized Convolutional Neural Network (CNN) model. Then, we applied a Layer-wise Relevance Propagation (LRP) technique, along with a unique pixel-flipping perofrmance metric, to clearly explain why our customized CNN model achieves high detection accuracy. We further evaluate the explanations generated by contemporary explainable AI methods such as Local Interpretable Model Agnostic Explanatio (LIME), Deep Taylor Decomposition (DTD), and Guided Backpropagation (GB) based on extensive computer-based simulations using a publicly available chest X-ray dataset.

The remainder of the paper is organized as follows. Section II surveys the relevant research work. Section III gives brief description about explainable AI methods we incorporated. Our customized image classification model with explainable AI is presented in section IV. The performance of our proposal is evaluated in section V. Finally, the paper is concluded in section VI.

II. Related Work

There are many different AI models have been proposed for COVID-19 detection and other lung diseases.

Wang et al. [8] proposed a COVID-Net architecture to find the abnormalities in chest X-ray images. It first projects the input features into lower dimensions. Then depth wise convolutions are performed to learn the spatial characteristics and preserve representational capacity. In the last step the features are converted to lower dimensions and then final features are produced by extending channel dimensionality. Any image preprocessing techniques or segmentation were not applied and the dataset was not balanced properly. To explain their model they highlighted the regions affected by COVID using 'GDInquire' method but was not well specific to critical areas.

Ghoshal *et al.* [9] estimated the uncertainty in deep learning solutions by using drop weights based Bayesian Convolutional Neural Networks (BCNN). This method was implemented to improve the classification results used in open source COVID-19 dataset and find the uncertainty in the prediction. They provided interpretability using attention maps based on BCNN approach and got some gripping results but there was not proper image preprocessing steps. They did not handle the imbalance in the data as the number if COVID-19 instances was 68 and the rest labels were around 1000 instances. They got performance of 89% but attention maps results were not so precise.

Karim *et al.* [10] made a deep COVID-19 explainer system using the ensemble model to extract features. They performed image preprocessing steps for enhancing the region of interest such as histogram equalization, filtering and non-sharp masking of chest X-ray images. In total, 15000 images were used having normal, pneumonia and COVID-19 instances. For the training, they used DenseNet, ResNet, and VGG transfer learning models [10]. They combined those models into ensemble using softmax class posterior averaging and prediction maximization for the best performance. They further explained the model using gradient guided class activation maps and LRP to identify the affected regions in lungs. Critical regions showed by the interpretability models were indefinite and also did not evaluate the quality of explanations generated.

Lucas *et al.* [11] performed segmentation and used three different architectures, referred to as VGG, DenseNet and Inception, to train their model. They further explained the model's predictions using LIME and Gradient-weighted Class Activation Mapping (Grad-CAM). While the LIME-based explanations were clear, the Grad-CAM-based explanation appeared vague. Furthermore, the work did not provide any explainability measure.

While the aforementioned research work attempted to interpret the performances of the underlying AI models, formulating a systematic, logical explanation for the prediction was overlooked, let alone evaluating the quality of those explanations considering the limited amount of data.

III. Explainable AI Methods Description

1) Layerwise Relevance Propagation (LRP): LRP is an attribute based back propagation method developed for structured neural networks considering the input are images or videos. In the first step standard forward pass is performed on the network and activations of every layer are stored. The relevance score predicted by the network in forward pass is propagated back layer by layer in the network till the input layer has reached. LRP follows $\alpha\beta$ rule where the value of α and β determines positive and negative relevance more the value of β more the negative relevance will be shown [12].In this project we ignored the bias and used alpha = 1 and beta = 0. To get only positive relevance of the class. LRP is used to find the relevant features of audio source localization, explaining EEG signal patterns and in microscopy to point the relevant cell structure and many more [13].

A. Deep Taylor Decomposition (DTD)

There are many layers in a deep neural network and every layer has a set of neurons. To reduce the total output error, neural network is trained by set of parameters at every layer. Thus, a structure of learned function is developed as a result of training a network. For instance, a neuron from the initial layer can get highly activated by considering a particular pixel. This neuron activation further can be used in higher layers to create complex non linearities having larger number of pixels. Deep Taylor decomposition (DTD) is inspired by divide and conquer concept. Consider one high layer neuron is $\mathrm{x_j}$ having relevance is $\mathrm{R_j}$ and its adjacent lower layer neuron is $\mathrm{x_i}$.We can decompose $\mathrm{R_j}$ to neurons in lower layer like $\mathrm{x_i}$ connected to $\mathrm{x_j}$.It can be assumed that $\mathrm{R_j(X_j)}$ are functionally related, we apply Taylor decomposition on this local function to redistribute the relevance R into the lower layer relevance $\mathrm{R_j}$ and so on. The first order terms give the relevance should be distributed on the neurons of lower layer. When applying Taylor decomposition on these sub functions reference points should be easier to find [14]. Thus breaking the complex function into sub function and applying this redistribution procedure in backward pass gives to pixel wise relevance which forms heatmap.

B. Guided Back propagation (GB)

Generally during the forward pass the relu activation sets the gradient which are negative to zero and while backpropagating both positive and negative gradients are passed. Thus, only in forward pass gradients are set to zero. Guided Backpropagation shows the key features of image that neurons detects. It sets negative gradient to zero while propagating forward and backward.

C. Local Interpretable Model Agnostic Explanation (LIME)

LIME is a method developed by [15] to explain the prediction of the model and it can be used to explain text, images and tabular data. The main idea behind lime is that it can explain any classification model and gives the understanding of the behavior of the model. Furthermore, it provides explanations by estimating local linear behavior of the model. To explain a prediction lime follows these steps:

1. Input data permutation: To explain prediction of particular image several artificial samples of that image is generated by turning super-pixel on and off. In this project we generated 1000 artificial samples of each images.

2. Predict the class: The class of every artificial sample is predicted.

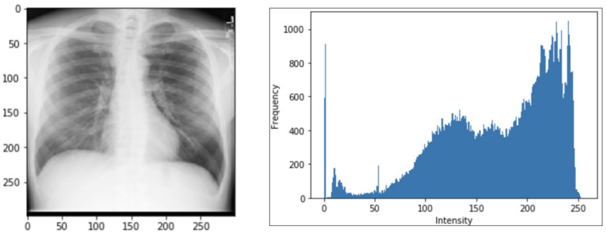


Fig. 1: Original Image and Histogram

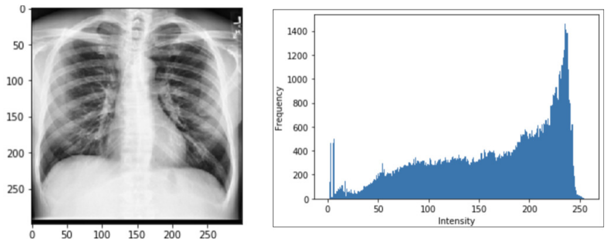


Fig. 2: Contrast Limited Adaptive Histogram

3. In this step, weights are calculated using cosine distance metric for every artificial sample. It finds out the distance between every perturbed sample and original input samples. Further the distance value obtained is mapped between zero to one with a kernel function. If the mapped value is closer to one the bigger the weight and its importance of that perturbed sample.
4. Now that we have artificial samples, weights and classes we fit it on a linear regression model. We sort the fitted coefficient and we get top 20 features of the image which played an important role in predicting the class.

IV. Proposed Lung Image Classification Model with Explainable AI

In this section, we adopt the use-case of lung disease prediction and develop a customized image classification model for this use-case. In particular, we aim to identify the lung region from every image using the lungs-finder library in Python and then perform image preprocessing and segmentation. Then, the processed image data are passed to our customized CNN model for lung disease classification.

A. Image Pre-processing

1) Adaptive Histogram Equalization: X-ray images are primarily used for diagnosis, even though other methods such as Magnetic Resonance Imaging (MRI) and Computer Tomography (CT) might provide more fine-grained information. The original form of imaging data lacks appropriate and precise quality because of lighting, noise, and contrast issues. The digital image processing approaches for enhancement have become significant for obtaining evidences from the images structure and effectively analyzing it for accurate diagnosis [16]. To improve the appearance of lung images, we perform the Contrast-Limited Adaptive Histogram Equalization (CLAHE). Traditionally, histogram equalization is applied on the whole image that sometimes results in loss of important information. To avoid this loss, CLAHE is applied on selected regions. CLAHE first divides the image into small parts and then perform histogram equalization is performed on every part in such a manner that the spread of intensity values is minimized.

As seen from Fig. 1 shows the original image histogram where the intensity values are distributed. In Fig. 2 adaptive method is shown which computes different histogram on every part of image and uses them to redistribute the lightness values of the image. As we can see that local contrast in the image is improved and edges of each region of image are enhanced.

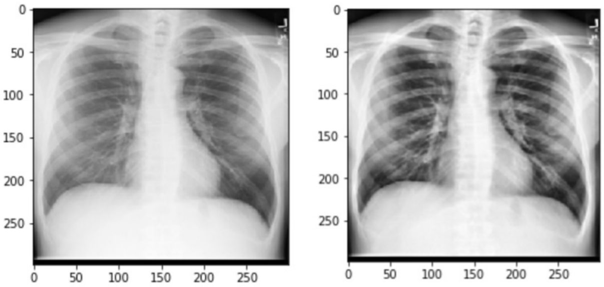


Fig. 3: (Left): Original Image, (Right): Anistropic Diffusion Image

2) Anisotropic Diffusion: Next, Anisotropic Diffusion Filter (ADF) also known as Perona-Malik Filtering is applied to the images to adaptively remove the noise, thereby maintaining the image edges. Thus, this technique aims at reducing image noise without removing significant parts of the image content, typically edges, lines or other details that are important for the interpretation of the image. It generates a group of parameterized images, and every output image is a combination of the original image and a filter depending on the local content of the original image [17].

The Fig 3.(Right) shows that the contrast in image is improved by applying CLAHE and on that image anisotropic diffusion is applied to remove noise and without blurring the edges.

B. Segmentation

Medical image processing is expected to have a class assigned to each pixel.Thus, every pixel should be associated with every class. For this purpose, the U-Net neural network, a specific CNN architecture [18], is utilized for a fast and precise segmentation of images. The convolution layers in U-Net are formed in top-down and bottom-up directions forming a U-shape network. The top-down direction is called the contracting section while the bottom-up direction is referred to as the expansive section. In contraction, a significant information of images is captured. On the other hand, expansion is exploited to localize the region of interest. The contracting section has in total 4 parts, each of which consists of two convolution layers having ReLU (Rectified Linear Unit) activation function as well as a max pooling layer. The expansive section consists of

4 subsections. The first two parts comprises convolutional, concatenation, and upsampling layers. The last block has three convolution, upsampling, and concatenation layers. The last part has dropout and output layers. Binary accuracy and loss are employed as performance metrics. When the image is passed to the input layer, the convolution, non-linearity, and downsampling layers are initiated. These operations reduce the size of the image. Then, after concatenation of the corresponding layers, in contracting and expanding sections, the size of image is increased. Thus, the output layer generates the image with the segmented lung region. As shown in Fig. 4: the segmentation gives precise lung region.

C. Dataset Description and Customized CNN Model for Automated Lung Disease Classification with Explainable AI

We use the dataset provided by [19] and [20] having 3,616 COVID-19 positive cases, 10,192 normal, and 6,012 lung opacity X-rays. As there is an imbalance in the data, we manually trim the normal images to 3,696 and opacity images to 3,529 for training. After getting segmented lungs, we divide the dataset for training and testing. For training, 3,375 COVID-19 samples, 3,695 normal samples, and 3,529 opacity samples are considered. Also, data augmentation is carried out by considering a zoom of 0.05, shear-range of 0.1, and the horizontal flip set to true. Based on the processed dataset, a DenseNet-201 CNN model [21] is trained by initializing weights to random and training all the layers. We train the model by setting the parameters mentioned in Table I on Google Colab Pro having specifications of 25 GB of Random Access Memory (RAM), high-end Graphics Processing Unit (GPU) P100, and long run time hours.

Next, to increase the model reliability and prediction quality, we train the model on segmented images. We apply the U-net learning algorithm to perform lung segmentation, taking chest X-ray images as input and giving a binary mask as output indicating the region of interest (ROI). Note that the U-Net training requires binary masks; however, the employed dataset does not have binary masks. Hence, we use chest X-ray dataset [22] that has 1000 binary masked lung X-rays to train the U-Net model and predicted binary masks for our dataset. The segmented lungs are trained on different deep neural networks with the parameters mentioned in Table I and the best performed model's predictions are explained. We employ the iNNvestigate tool [23] to explain the model's test data predictions with Layer-wise Relevance Propagation (LRP), Deep Taylor Decomposition (DTD), and Guided Backpropagation (GB). We also explain the predictions using a model-agnostic method called LIME (Local Interpretable Model-Agnostic Explanations) developed by researchers in [14] to explain the behavior of the complex model by providing the top features of the image. The generated explanation's complexity is calculated by image entropy. Furthermore, using the pixel flipping algorithm, the quality of heatmaps are evaluated.

TABLE I: Training parameters.

Parameters	Values
Training data	10,596
Testing data	840
Epochs	50
Optimizer	Adam (lr= 0.0001)
Loss	Categorical Cross-entropy
Layers trainable	True
Weights	Randomly initialized

V. Performance Evaluation

In this section, we evaluate both the accuracy of the adopted classification model and its explainability performance. In addition to our customized CNN model using DenseNet-201 architecture, the segmented dataset is used to further train DenseNet-121, ResNet-50, InceptionV3 models [21] for comparative analytics. The test dataset comprised a total of 840 images with 216 COVID-19 images, 282 normal images, and 342 lung opacity images. Among these methods, DenseNet-201 demonstrated good prediction performance. It can be observed from the confusion matrix Fig. 6 that the model could classify majority of the X-rays.

TABLE II: F1 scores and accuracies of different classification models.

Model	F1 (COVID-19)	F1 (Opacity)	F1(Normal)	Accuracy
ResNet50	69	83	83	79
InceptionV3	62	81	56	68
DenseNet-121	78	80	81	80
DenseNet-201	84	84	83	84

It can be observed from Table II that DenseNet-201 achieved the highest F1 score of COVID-19, opacity, and normal class followed by our adopted DenseNet-121 model. In the next step, we picked random images from the test dataset and explained the prediction by implementing the above various explaiable AI methods. In Fig. 7, generated heatmaps using LIME, DTD, GB, and LRP demonstrate the pixels that played an important role in prediction. LIME was able to exhibit the top 20 features. In the GB method, the explanation demonstrates the features that the neuron unit detects causing the pixels to spread in an equal intensity in the image. In the DTD method, the explanation is more transparent as it focuses on one region of the image, i.e., it attempts to highlight the white matter in lung opacity and COVID-19 images. When it comes to LRP, the explanation is almost similar to that of the DTD method because the pixels continue to become darker in the region where there are opacities. If the X-ray is normal, the explanation does not highlight a specific region. Thus, these methods explain by going through the model and gives the pixel importance. However, the LIME method explained the model behavior by generating the artificial points and fitting them in a linear model. Thus, it provided the top features of the image, which plays an essential role in getting this prediction. Even though LIME is able to explain any machine learning model, the result in Fig. 7 demonstrates that it is unstable, i.e., it creates some attributions irrelevant to human explanations.

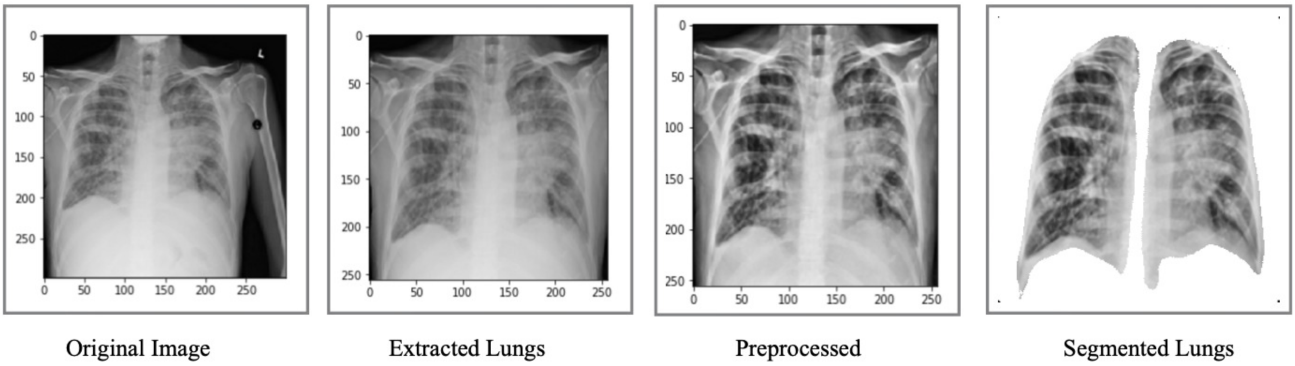


Fig. 4: Image Preprocessing

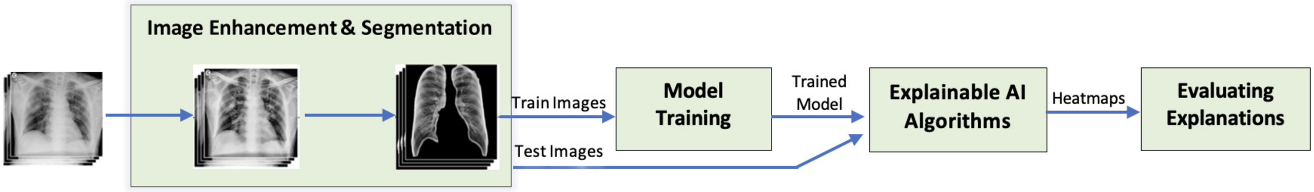
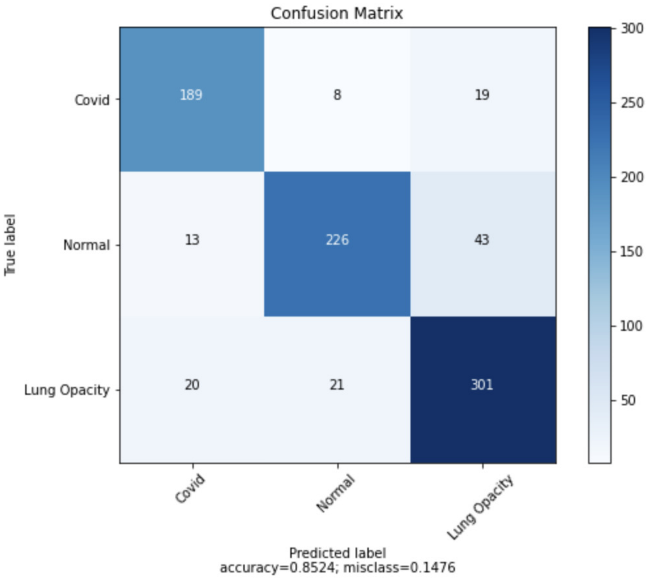


Fig. 5: Proposed automated lung disease disorder classification system with explainable AI.



True label \ Predicted label	Covid	Normal	Lung Opacity
Covid	189	8	19
Normal	13	226	43
Lung Opacity	20	21	301

Fig. 6: Confusion Matrix

To evaluate the complexity of explanations, we calculated the image entropy, which measures the randomness in an image. If a heatmap contains only relevant regions, then it is considered to have low randomness and can be compressed easily due to small size. On the other hand, when a heatmap contains more than relevant information, it is attributed with a high randomness and will be large in size. Thus, the explainable AI method having low randomness focuses on significant information in an image. We took the average of the entropy values of 30 explanations. Our experiments revealed that the GB method demonstrated the highest average value of 4.0, i.e., a very high randomness in explanation followed by LIME with 3.7 where the values are distributed over a large range. The DTD method exhibited the lowest complexity with 3.35, i.e., the least randomness. The complexity of LRP is also comparable with an average value of 3.4.

Next, we omit the LIME method due to its inability to offer cohesive explainability, and further evaluate the explainability performances of LRP, DTD and GB methods by using the pixel flipping performance metric on ten test images. As seen from Fig. 8, DTD and LRP methods drop more steeply compared to GB. Thus, the explanations generated by these methods can be considered more precise and demonstrate relevant areas in the image.

By observing the image entropy and pixel flipping results, it is clear that DTD offers slightly better results than LRP. This happens because both of these methods redistribute relevance. However, the DTD method does not consider the whole network function due to its divide and conquer approach for redistributing relevances. On the other hand, LRP considers the whole network function, which is why it can be regarded as the most viable explainable method to account for the reasonably accurate prediction performance of the underlying lung image classification model.

VI. Conclusion

In this paper, we incorporated different explainable AI methods to explain the performance of the underlying AI model for lung image classification. Considering the evaluation metrics to measure the quality of explanation by implementing image entropy and pixel-flipping algorithm, the Deep Taylor Decomposition (DTD) method showed the best performance. However, it was observed that Layer-wise Relevance Propagation (LRP) provided almost similar results. In the future, these methods should be evaluated on

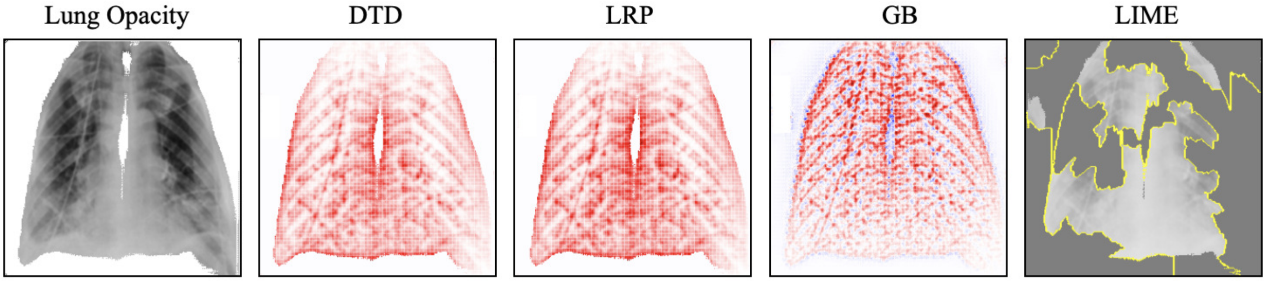


Fig. 7: Comparison of explainable AI methods on the reigons-of-interest in an example chest X-ray image. Note that for brevity only one example image is included here. Similar observations have been remarked for all other example images.

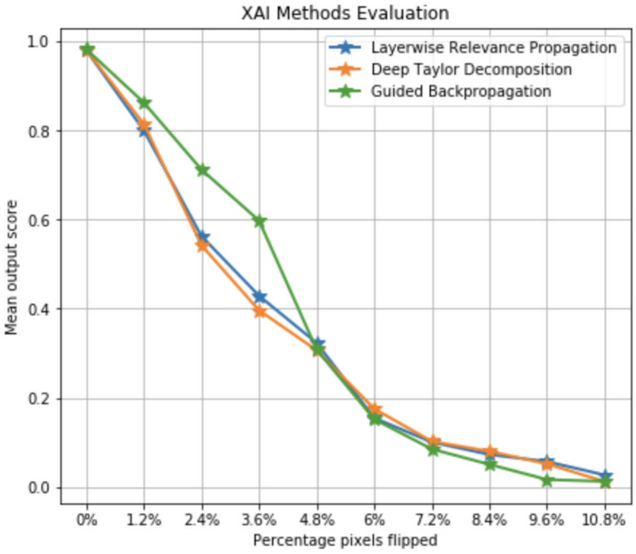


Fig. 8: Pixel-flipping graph.

other, more robust metrics to gain a deeper understanding of the underlying model's behavior. The heatmaps generated still consist of some irrelevance of bone structure. Thus, the bone suppression technique during image preprocessing can help improve the model's performance and can give a more precise explanation. Thus, adding explainability to the system can open doors for many AI healthcare treatments and can also help to gain new insights and improve current systems.

References

[1] M. Hammad, A. Maher, K. Wang, F. Jiang, and M. Amrani, "Detection of abnormal heart conditions based on characteristics of ECG signals," *Measurement*, vol. 125, pp. 634–644, 2018.

[2] H. Khalil *et al.*, "Classification of diabetic retinopathy types based on convolution neural network (CNN)," *Menoufia Journal of Electronic Engineering Research*, vol. 28, no. ICEEM2019-Special Issue, pp. 126–153, 2019.

[3] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[4] Z. Fadlullah, M. M. Fouda, A.-S. K. Pathan, N. Nasser, A. Benslimane, and Y.-D. Lin, "Smart IoT solutions for combating the COVID-19 pandemic," *IEEE Internet of Things Magazine*, vol. 3, no. 3, pp. 10–11, 2020.

[5] T. Tazrin, M. M. Fouda, Z. M. Fadlullah, and N. Nasser, "UV-CDS: An energy-efficient scheduling of UAVs for premises sterilization," *IEEE Transactions on Green Communications and Networking*, vol. 5, no. 3, pp. 1191–1201, 2021.

[6] S. Sakib, T. Tazrin, M. M. Fouda, Z. M. Fadlullah, and M. Guizani, "DL-CRC: Deep learning-based chest radiograph classification for COVID-19 detection: A novel approach," *IEEE Access*, vol. 8, pp. 171 575–171 589, 2020.

[7] S. Sakib, M. M. Fouda, Z. Md Fadlullah, and N. Nasser, "On COVID-19 prediction using asynchronous federated learning-based agile radiograph screening booths," in *ICC 2021 - IEEE International Conference on Communications*, 2021, pp. 1–6.

[8] L. Wang and A. Wong, "COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images," *Science Reports*, 2020.

[9] B. Ghoshal and A. Tucker, "Estimating uncertainty and interpretability in deep learning for coronavirus (COVID-19) detection," *arXiv preprint arXiv:2003.10769*, 2020.

[10] M. R. Karim *et al.*, "DeepCOVIDExplainer: Explainable COVID-19 diagnosis based on chest X-ray images," *arXiv preprint arXiv:2004.04582*, 2020.

[11] L. O. Teixeira *et al.*, "Impact of lung segmentation on the diagnosis and explanation of COVID-19 in chest X-ray images," *arXiv preprint arXiv:2009.09780*, 2021.

[12] G. Montavon, W. Samek, and K. R. Müller, "Methods for interpreting and understanding deep neural networks," *Digital Signal Processing*, vol. 73, no. 1-15, 2018.

[13] Samek, W. Montavon, G. Vedaldi, A.Hansen, L. K, and Muller, "Explainable AI: Interpreting, explaining and visualizing deep learning. lecture notes in computer science," *Digital Signal Processing*, 2019.

[14] G. Montavon, S. Lapuschkin, A. Binder, W. Samek, and K.-R. Müller, "Explaining nonlinear classification decisions with deep taylor decomposition," *Pattern Recognition*, vol. 65, pp. 211–222, 2017.

[15] M. T. Ribeiro, S. Singh, and C. Guestrin, ""why should i trust you?": Explaining the predictions of any classifier," 2016.

[16] S. J. Attia, "Enhancement of chest X-ray images for diagnosis purposes," *Journal of Natural Sciences Research*, vol. 6, no. 2, pp. 43–46, 2016.

[17] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990.

[18] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," *arXiv preprint arXiv:1505.04597*, 2015.

[19] M. E. H. Chowdhury *et al.*, "Can AI help in screening viral and COVID-19 pneumonia?" *IEEE Access*, vol. 8, pp. 132 665–132 676, 2020.

[20] T. Rahman *et al.*, "Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images," *Computers in Biology and Medicine*, vol. 132, article no. 104319, 2021.

[21] S.-H. Wang and Y.-D. Zhang, "DenseNet-201-based deep neural network with composite learning factor and precomputation for multiple sclerosis classification," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 16, no. 2s, article no. 60, Jun. 2020.

[22] L. Wu. Lung segmentation on RSNA pneumonia detection dataset. Last accessed: 9 May 2021. [Online]. Available: https://drive.google.com/drive/folders/1gISKPOiDuZTAXkGeQ6-TMb3190v4Xhyc

[23] M. Alber *et al.*, "iNNvestigate neural networks!" *arXiv preprint arXiv:1808.04260*, 2018.