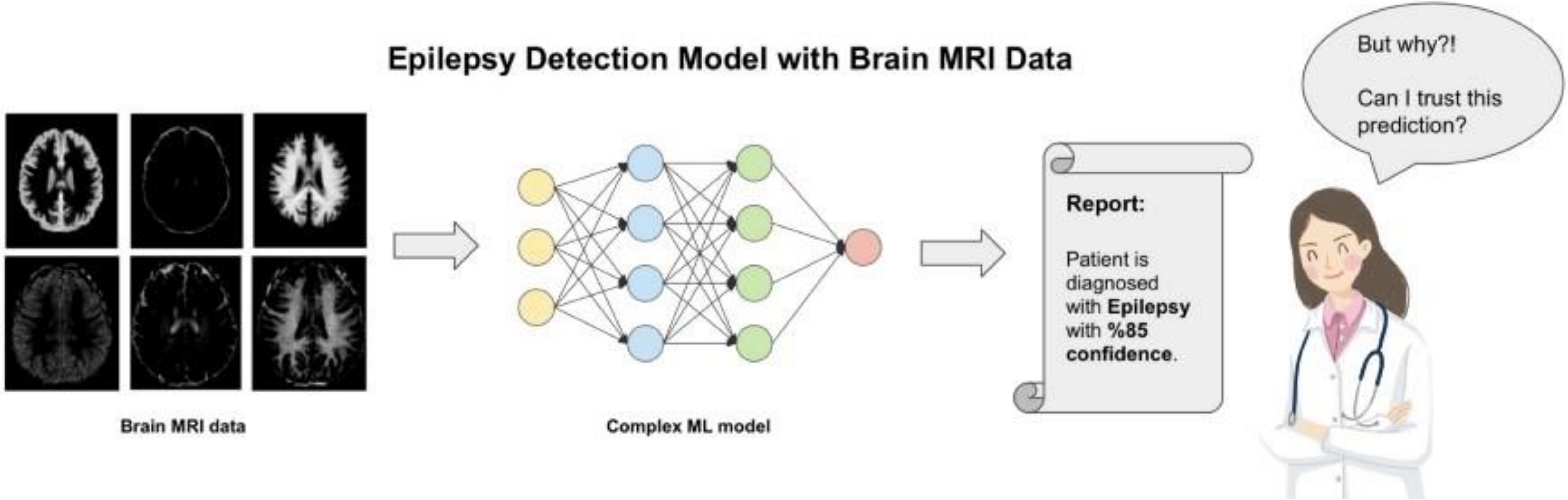
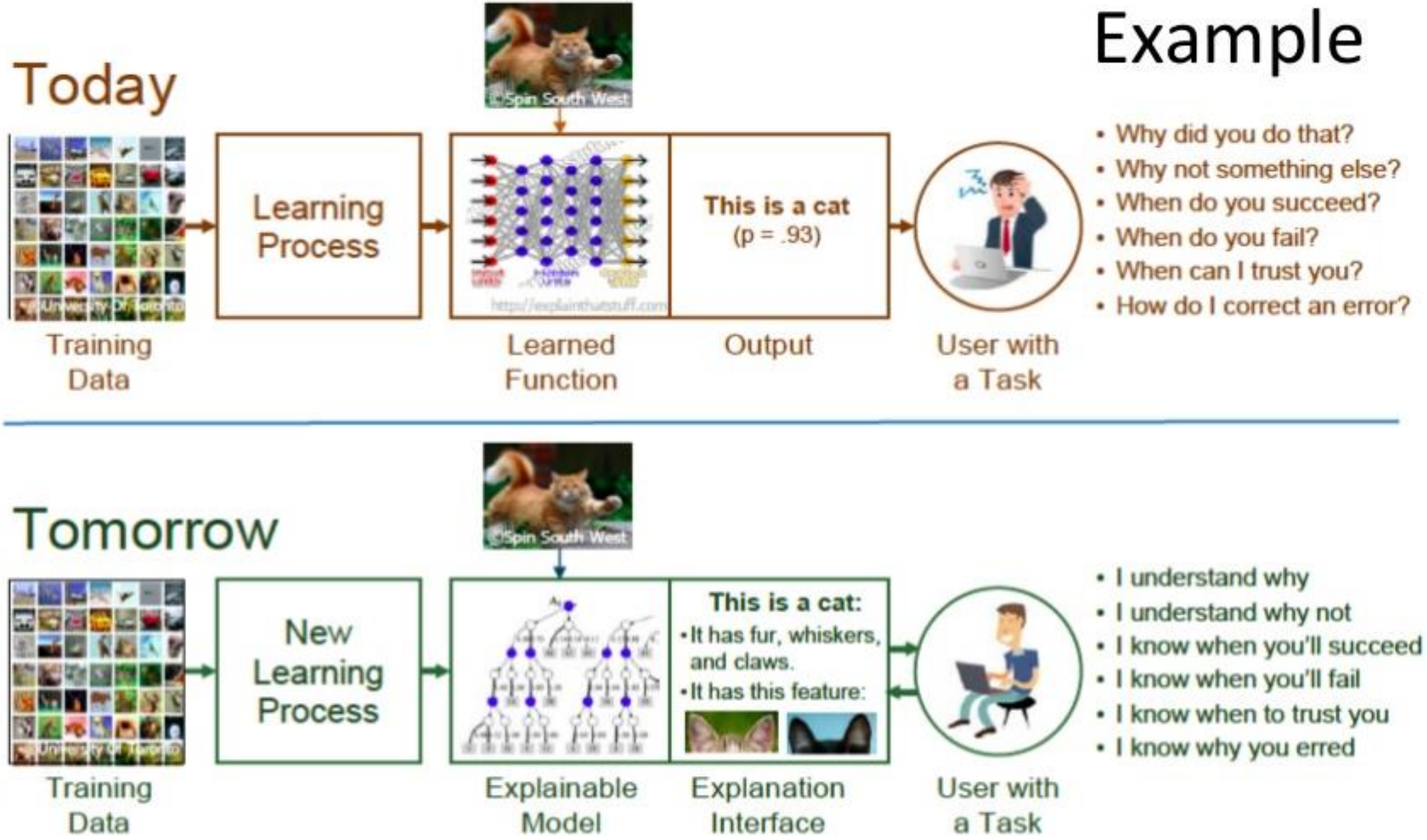
Wyjaśnialna sztuczna inteligencja

Weronika Hryniewska

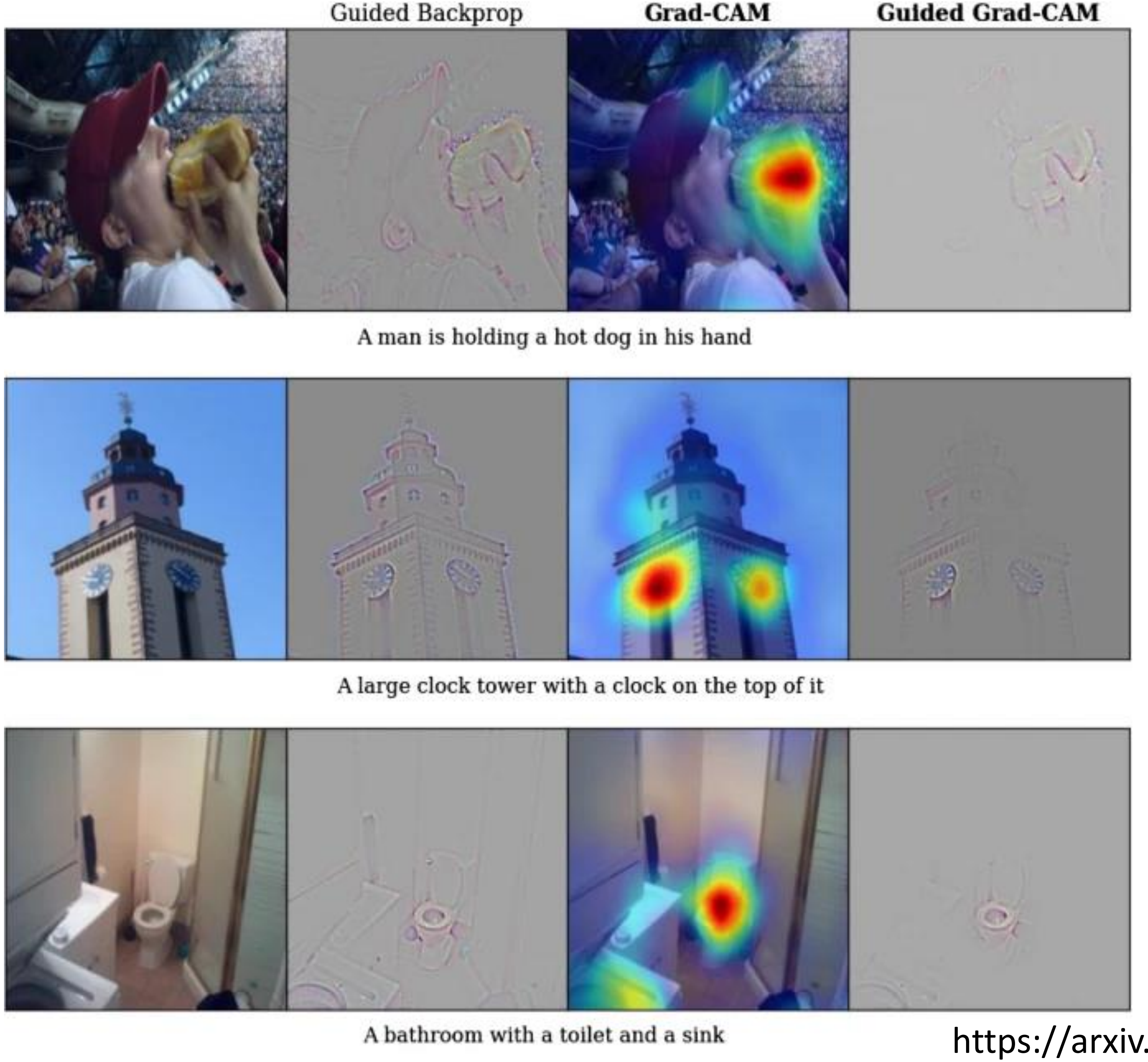


https://medium.com/altaml/how-explainable-artificial-intelligence-xai-can-help-us-trust-ai-8f01b574102d

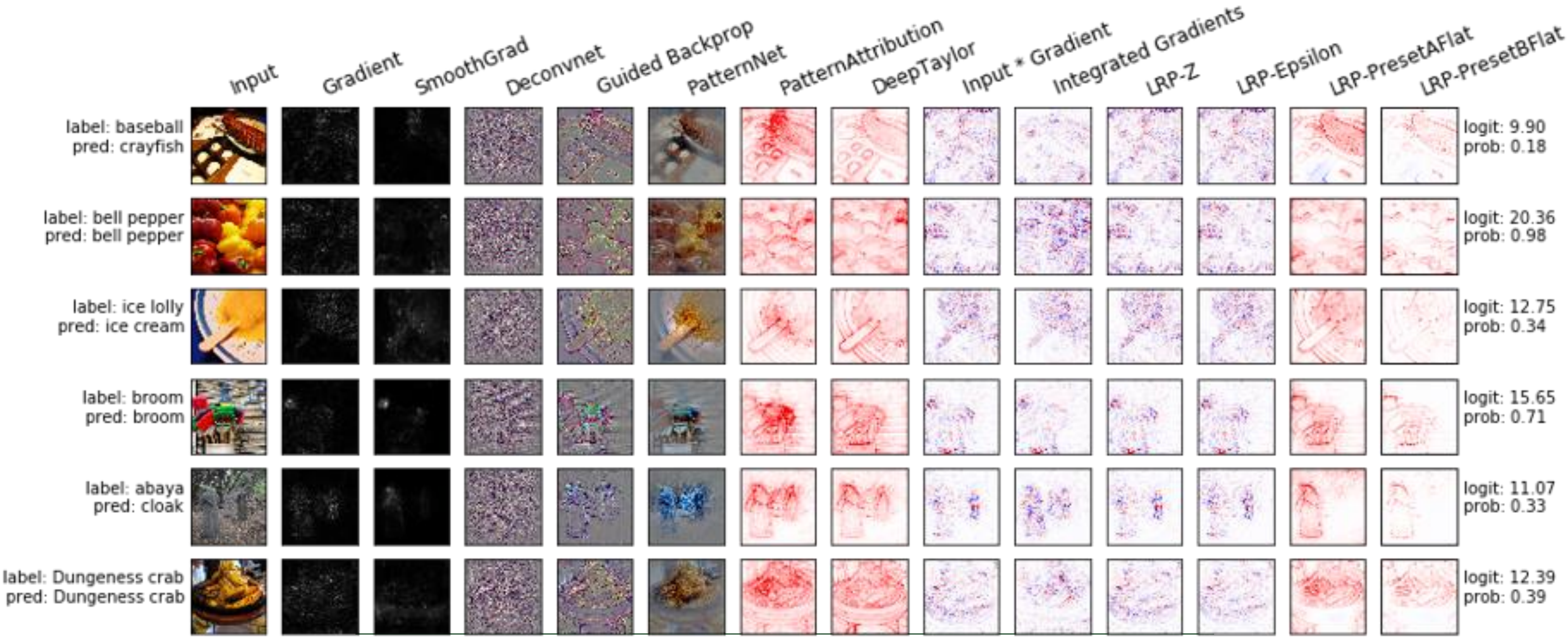
Example



https://www.slideshare.net/ManojkumarParmar/explainable-ai-xai

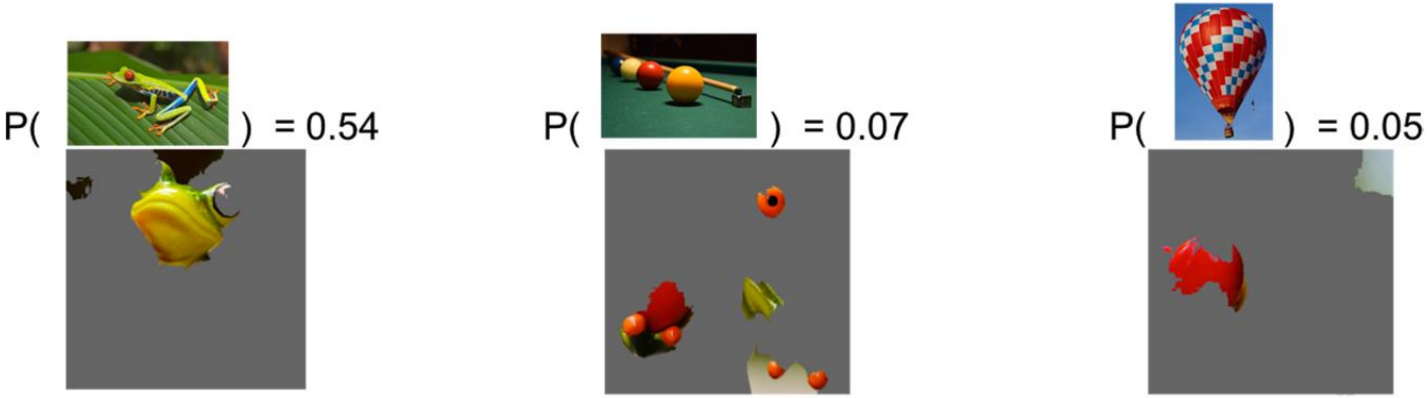


https://arxiv.org/pdf/1610.02391.pdf

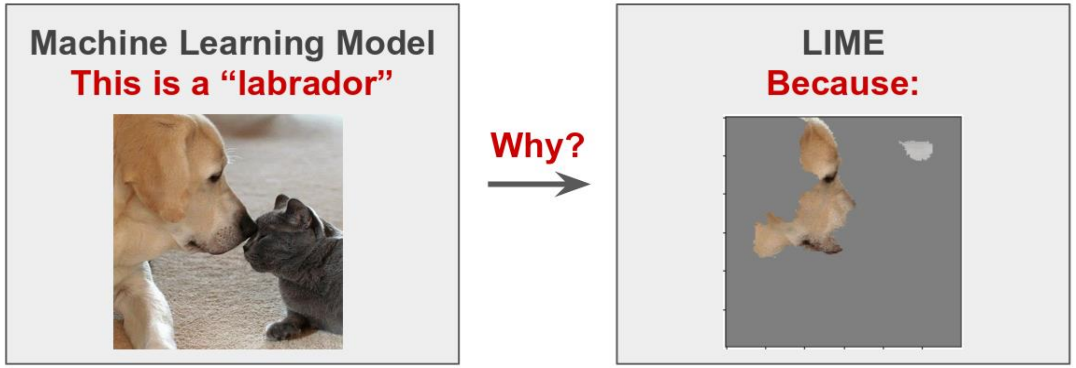


https://github.com/albermax/innvestigate

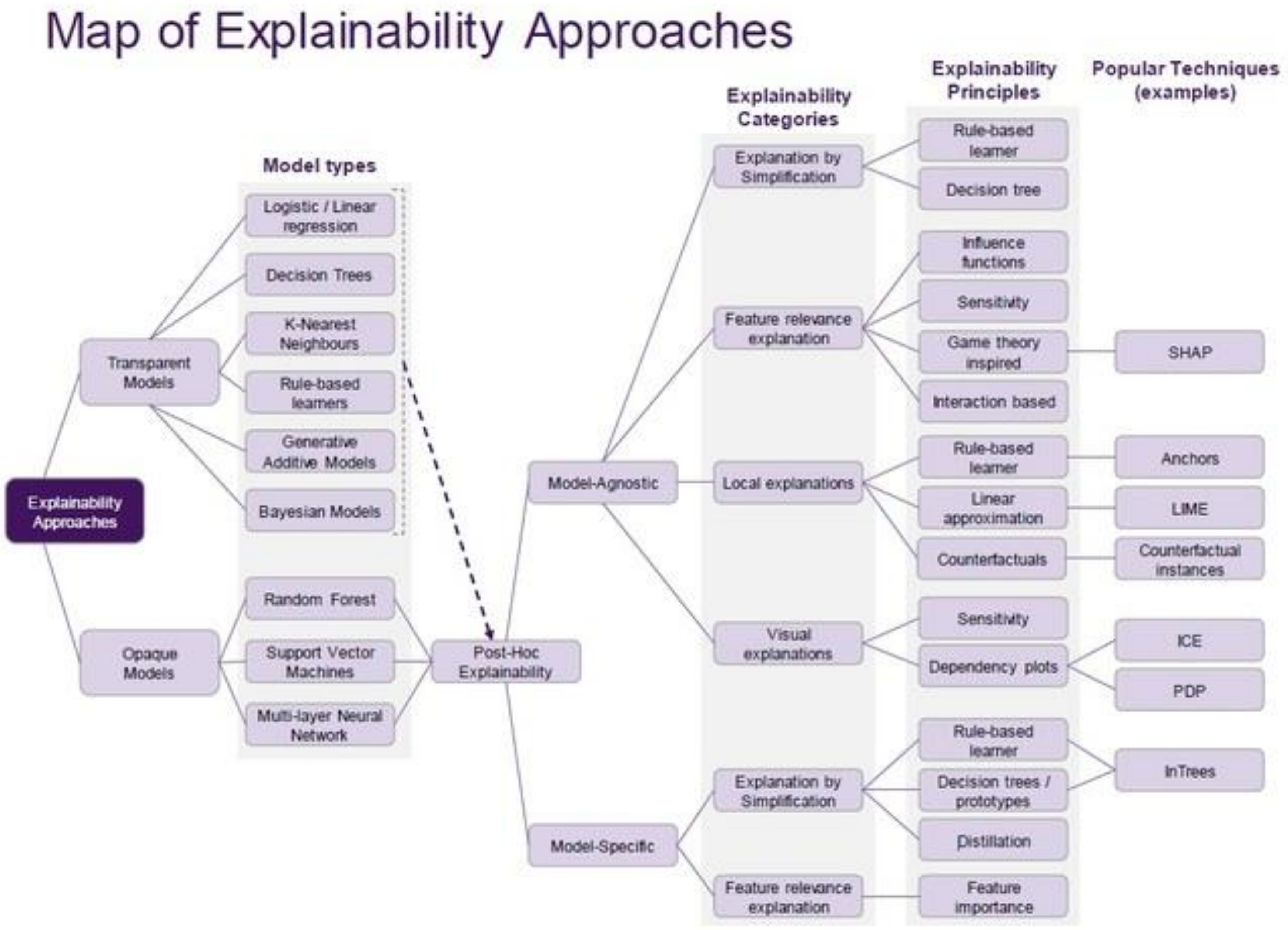
Local Interpretable Model-Agnostic Explanations (LIME)



www.oreilly.com/content/introduction-to-local-interpretable-model-agnostic-explanations-lime



https://towardsdatascience.com/interpretable-machine-learning-for-image-classification-with-lime-ea947e82ca13

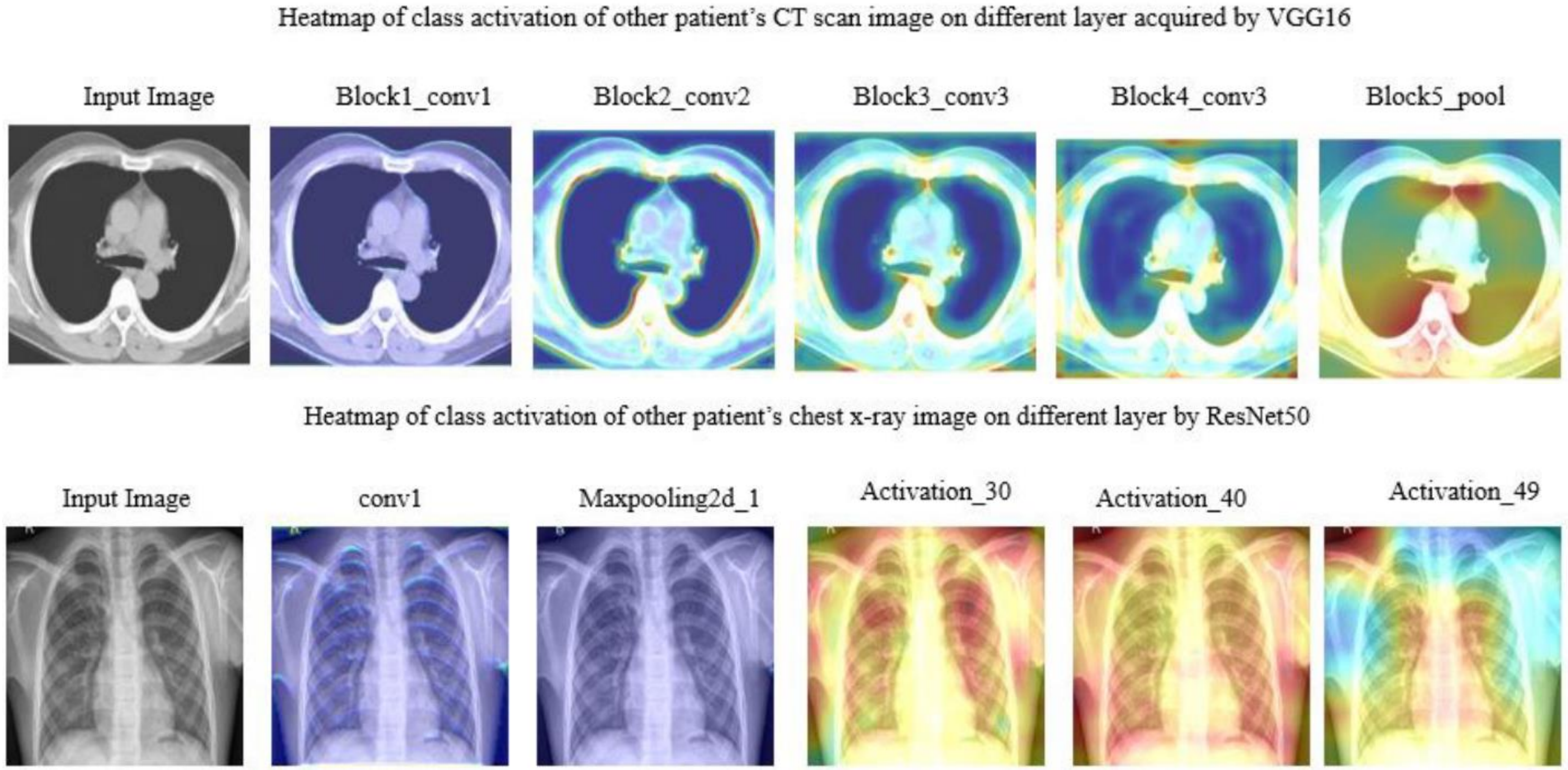


https://www.datasciencecentral.com/profiles/blogs/a-taxonomy-of-explainable-xai-ai-models

Przykłady metod XAI dla zdjęć płuc chorych na Covid-19

Weronika Hryniewska

Abbas, A., Abdelsamea, M. M., & Gaber, M. M. (2020). *Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network*. http://arxiv.org/abs/2003.13815



Brunese, L., Mercaldo, F., Reginelli, A., & Santone, A. (2020). Explainable Deep Learning for Pulmonary Disease and Coronavirus COVID-19 Detection from X-rays. *Computer Methods and Programs in Biomedicine*, *196*, 105608. https://doi.org/10.1016/j.cmpb.2020.105608

Are such distances from the marked areas acceptable?

With red arrow the radiologists marked the areas where the COVID-19 disease is manifested, as is shown from the activation maps, the proposed models rightly highlighted the areas marked by the radiologist.

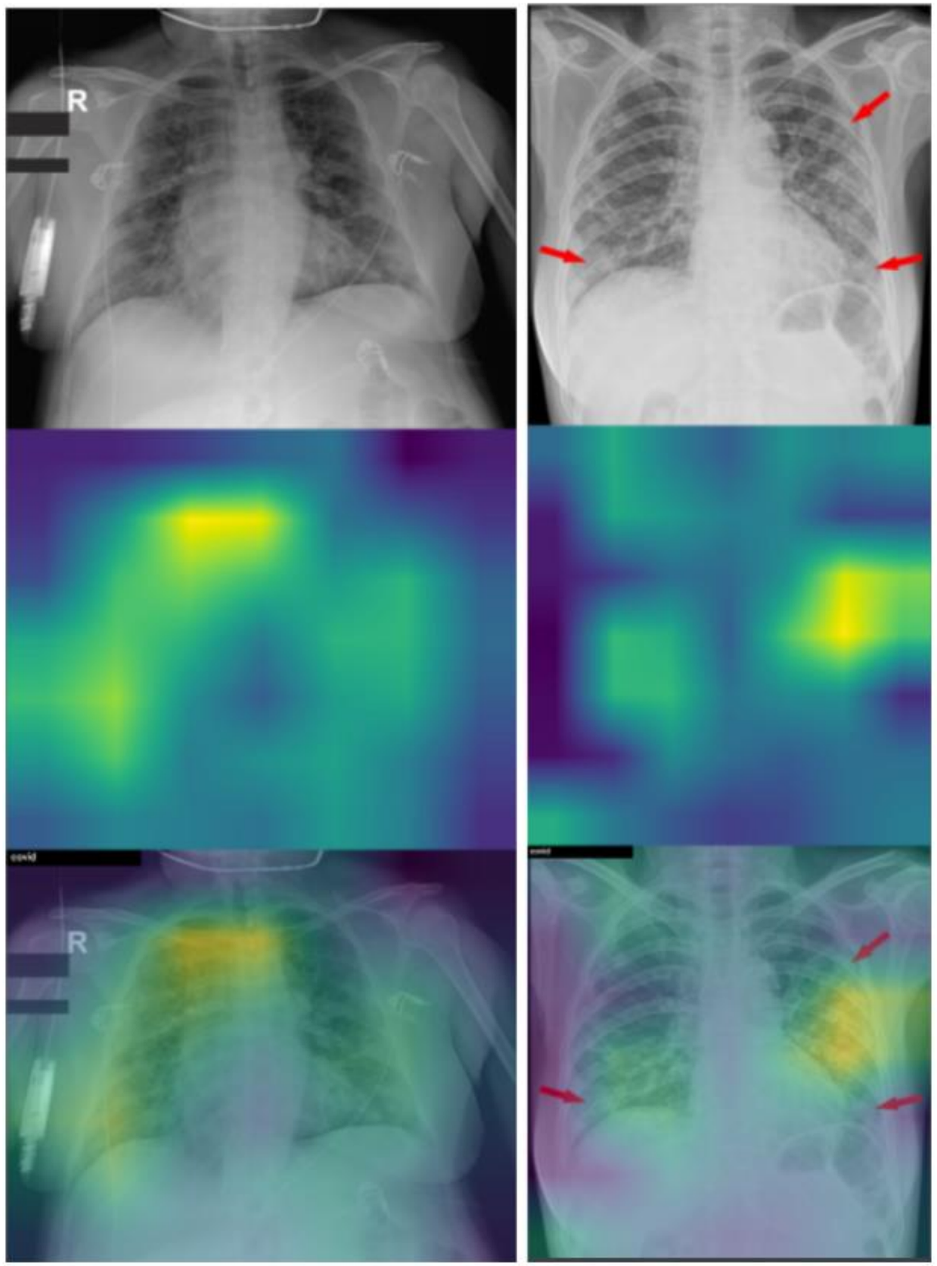
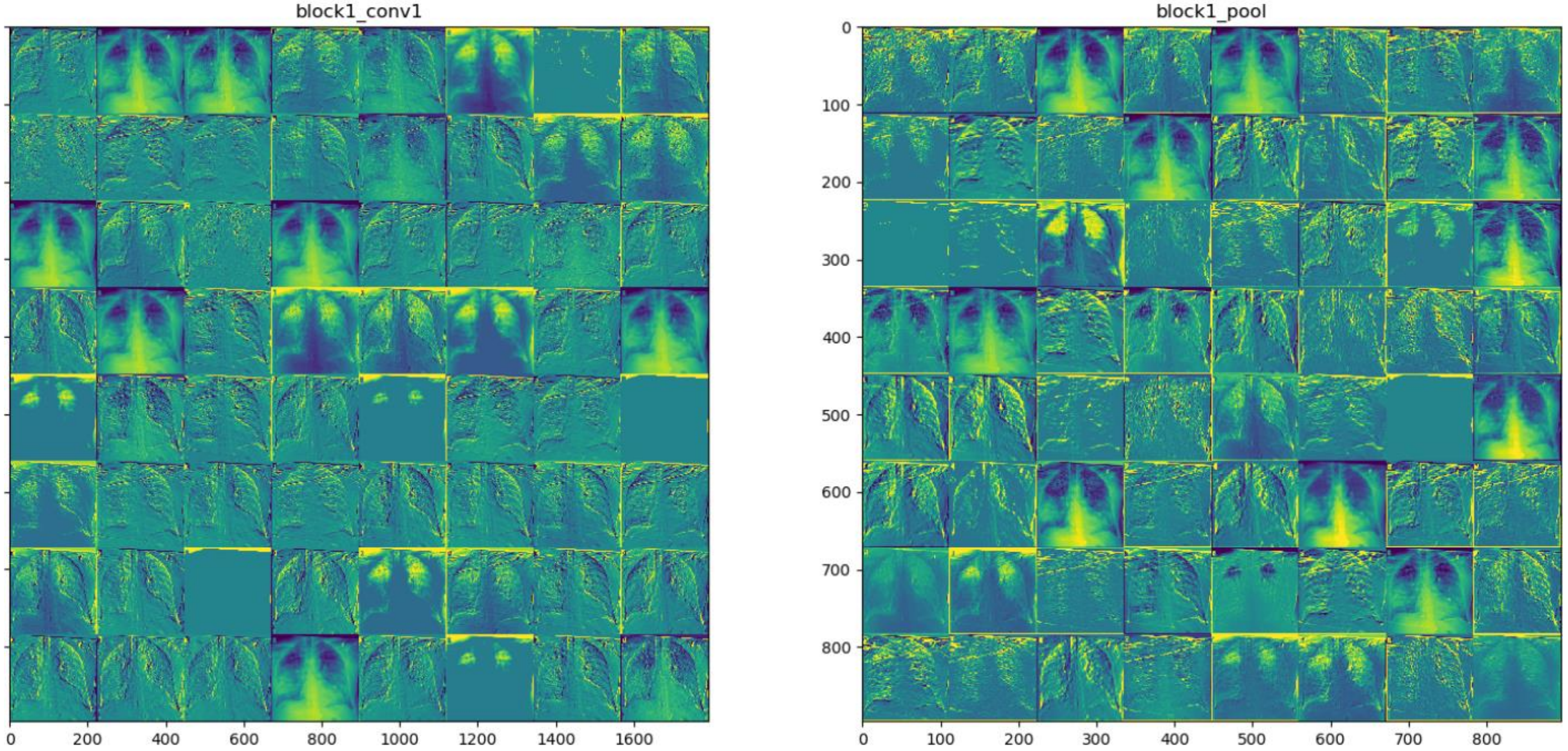


Fig. 9. Examples of COVID-19 model activation maps.

Sharma, V., & Dyreson, C. (2020). *COVID-19 detection using Residual Attention Network an Artificial Intelligence approach*. http://arxiv.org/abs/2006.16106



(a) Feature Map at Convolution Layer 1

(b) Feature Map at pooling Layer

Karim, M. R., Döhmen, T., Rebholz-Schuhmann, D., Decker, S., Cochez, M., & Beyan, O. (2020). *DeepCOVIDExplainer: Explainable COVID-19 Diagnosis Based on Chest X-ray Images*. http://arxiv.org/abs/2004.04582

Are such big areas relevant?

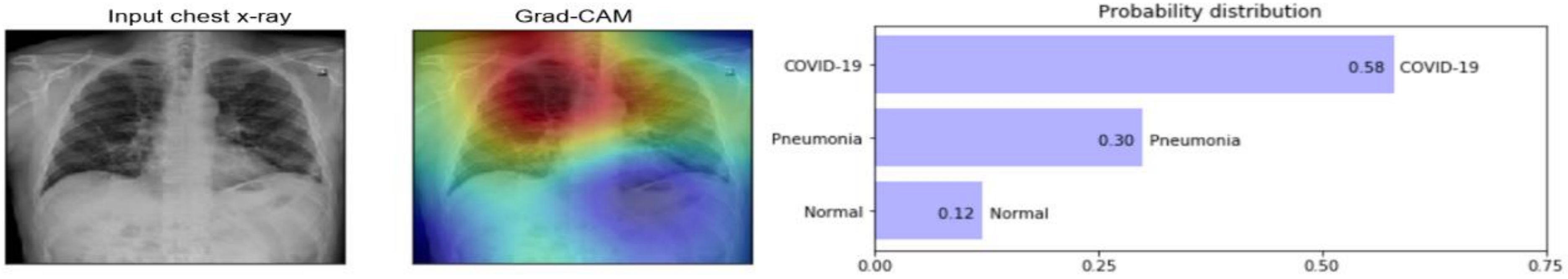


Figure 7: The input chest x-ray classification, decision visualization with Grad-CAM and explanation

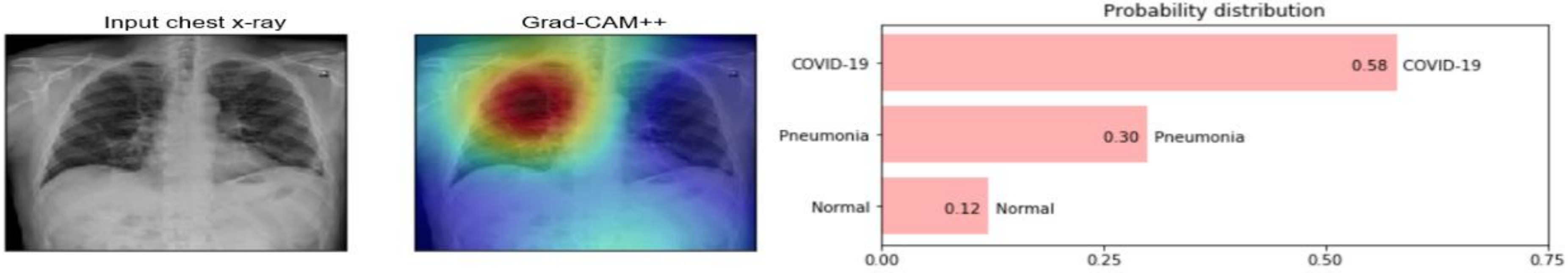


Figure 8: The input chest x-ray classification, decision visualization with Grad-CAM++ and explanation

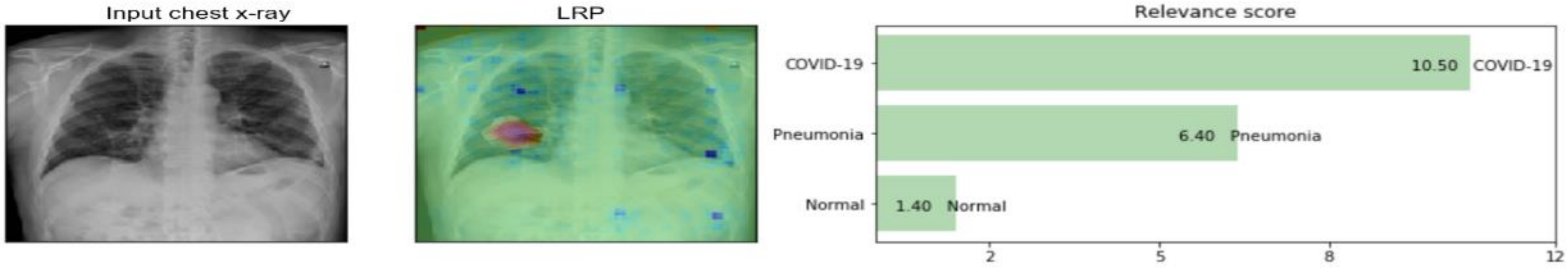
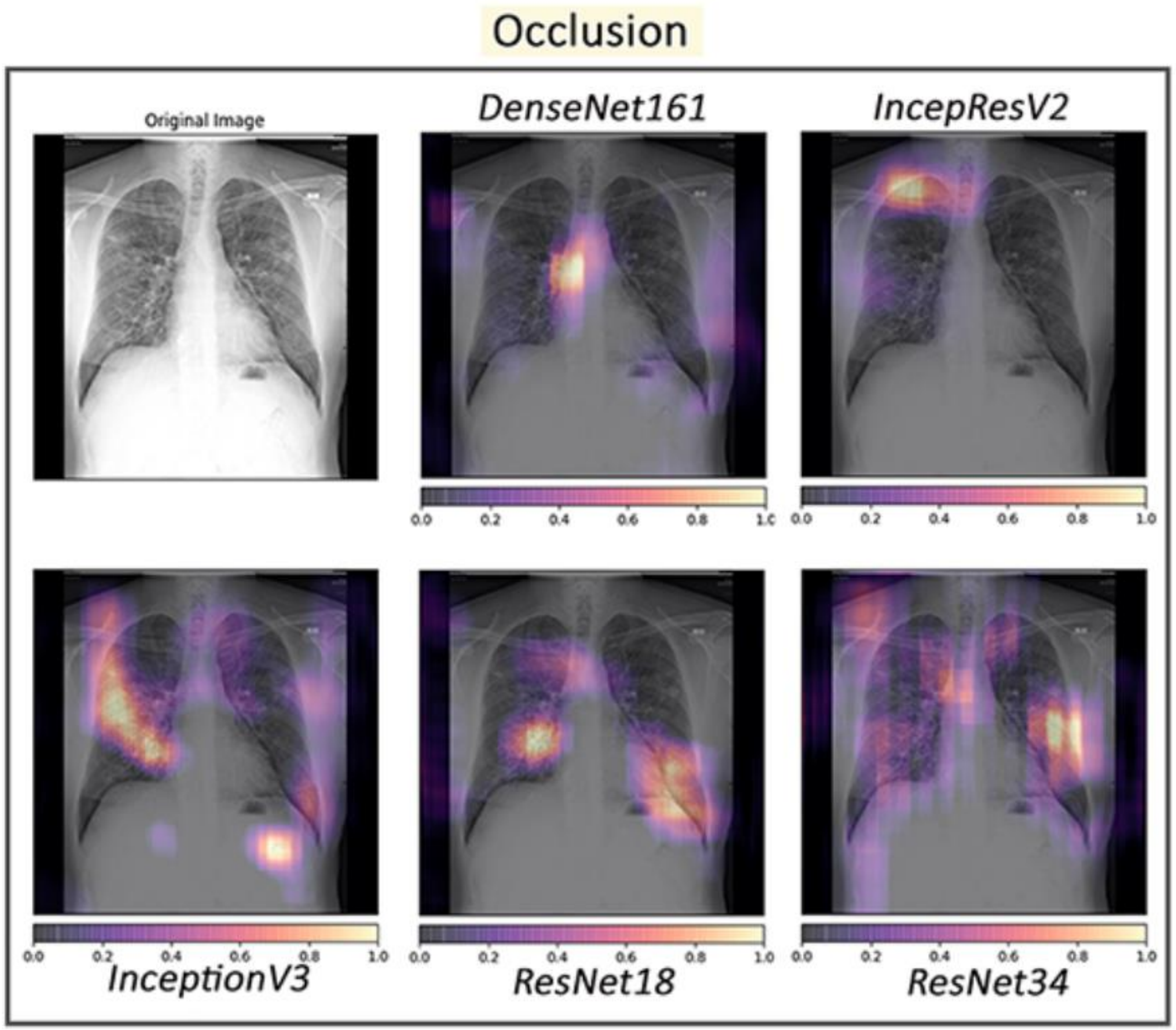
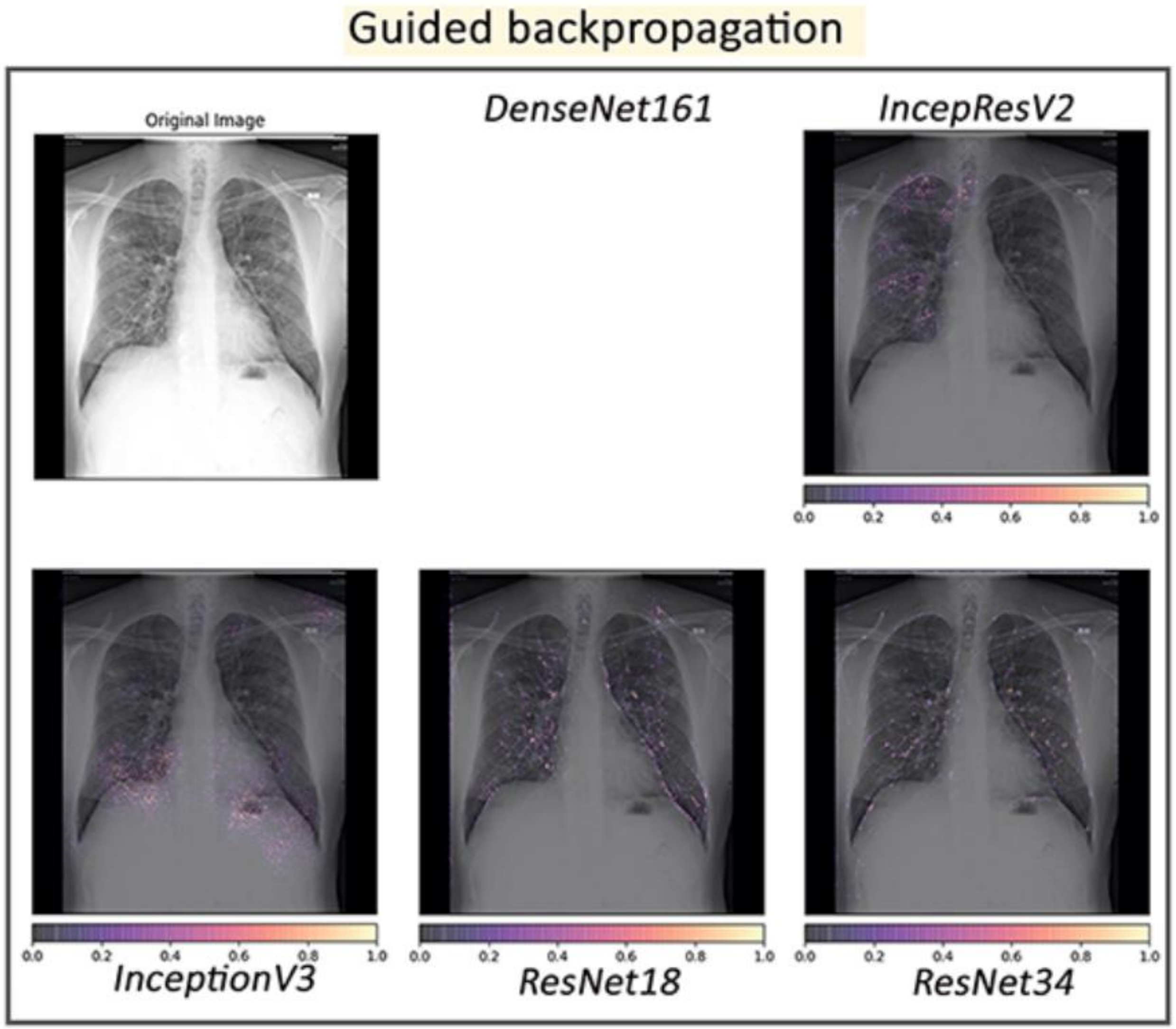


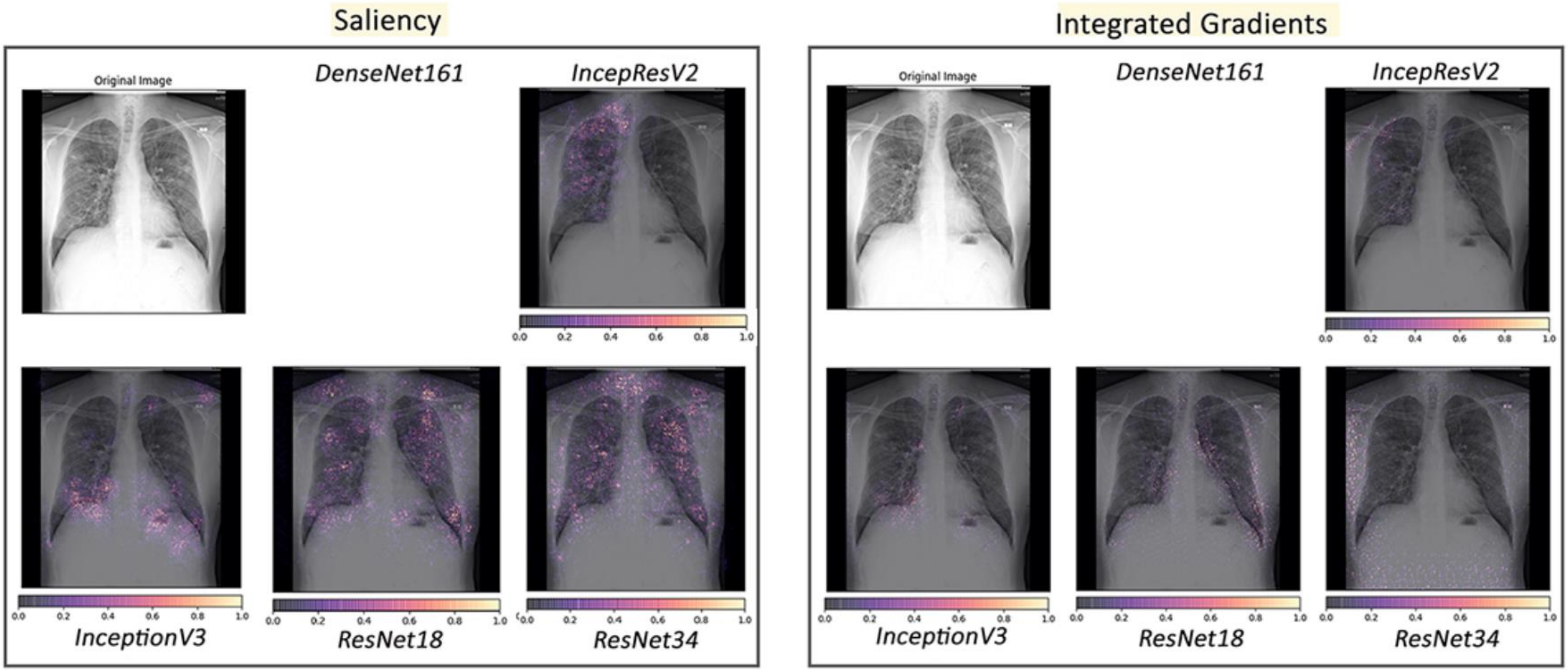
Figure 9: The input chest x-ray classification, decision visualization with LRP and explanation

Chatterjee, S., Saad, F., Sarasaen, C., Ghosh, S., Khatun, R., Radeva, P., Rose, G., Stober, S., Speck, O., & Nürnberger, A. (2020). *Exploration of Interpretability Techniques for Deep COVID-19 Classification using Chest X-ray Images*. http://arxiv.org/abs/2006.02570

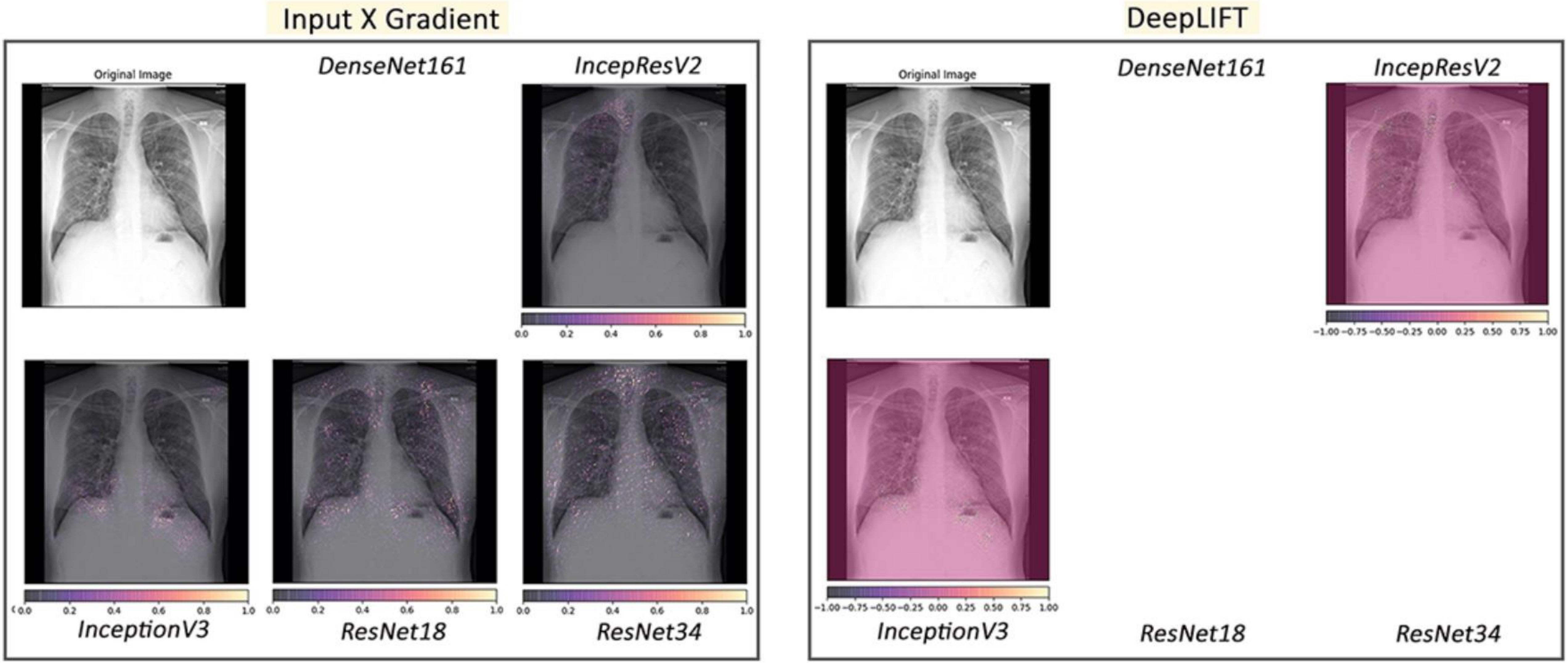
Which visualization method is most useful?





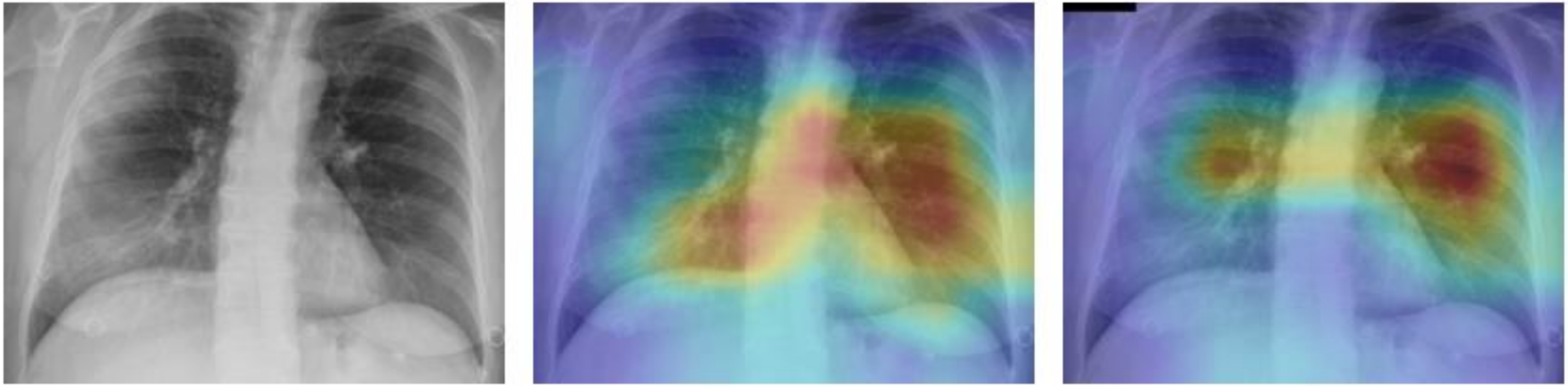


Saliency
Original Image
DenseNet161
IncepResV2
InceptionV3
ResNet18
ResNet34
Integrated Gradients
Original Image
DenseNet161
IncepResV2
InceptionV3
ResNet18
ResNet34
0.0
0.2
0.4
0.6
0.8
1.0

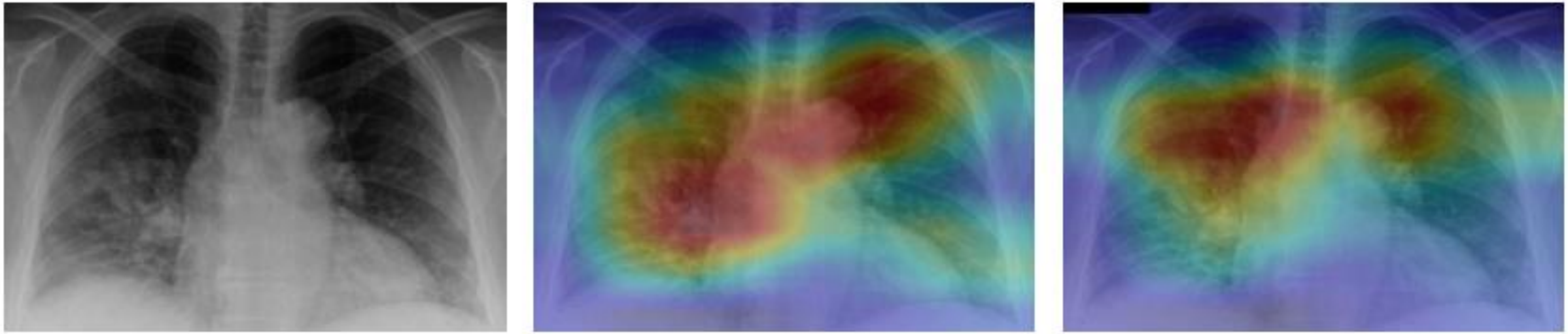

Input X Gradient
Original Image
DenseNet161
IncepResV2
0.0 0.2 0.4 0.6 0.8 1.0
InceptionV3
ResNet18
ResNet34
DeepLIFT
Original Image
DenseNet161
IncepResV2
−1.00 −0.75 −0.50 −0.25 0.00 0.25 0.50 0.75 1.00
InceptionV3
ResNet18
ResNet34

Signoroni, A., Savardi, M., Benini, S., Adami, N., Leonardi, R., Gibellini, P., Vaccher, F., Ravanelli, M., Borghesi, A., Maroldi, R., & Farina, D. (2020). *End-to-end learning for semiquantitative rating of COVID-19 severity on Chest X-rays*. 1–22. http://arxiv.org/abs/2006.04603

Is the area of changes related to the severity of the disease?



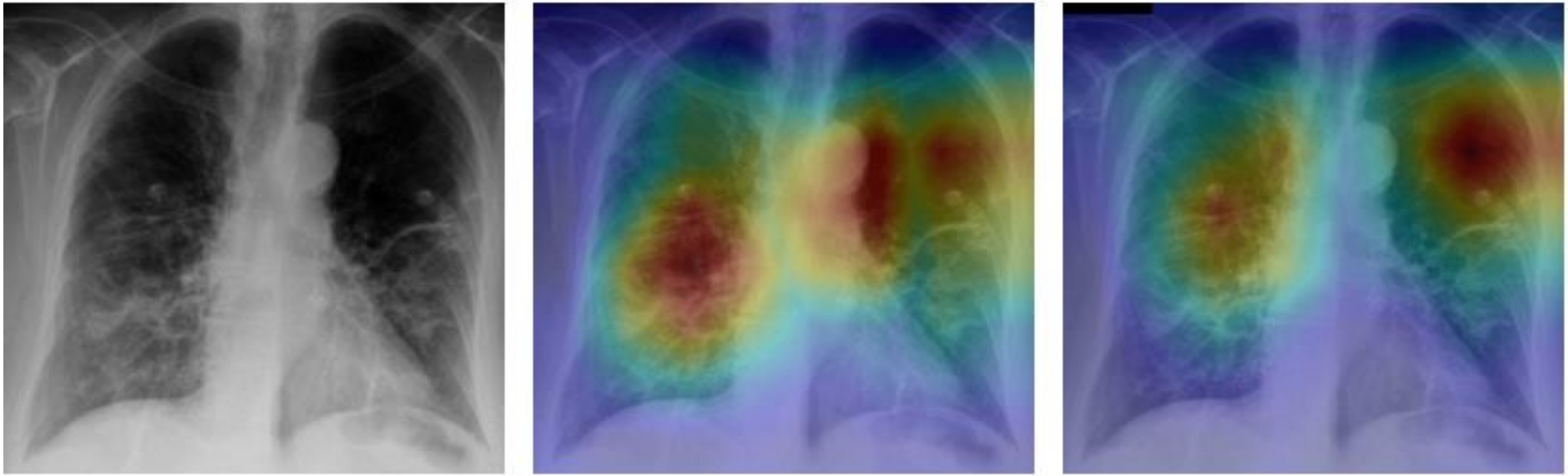
(a) Original Positive (Mild)　(b) why positive　(c) why negative



(a) Original Positive (Moderate)

(b) why positive

(c) why negative



(a) Original Positive (Severe) (b) why positive (c) why negative

(a) Original Negative (b) why positive (c) why negative

Ouyang, X., Huo, J., Xia, L., Shan, F., Liu, J., Mo, Z., Yan, F., Ding, Z., Yang, Q., Song, B., Shi, F., Yuan, H., Wei, Y., Cao, X., Gao, Y., Wu, D., Wang, Q., & Shen, D. (2020). Dual-Sampling Attention Network for Diagnosis of COVID-19 from Community Acquired Pneumonia. *IEEE Transactions on Medical Imaging*, *39*(XX), 1–1. https://doi.org/10.1109/tmi.2020.2995508

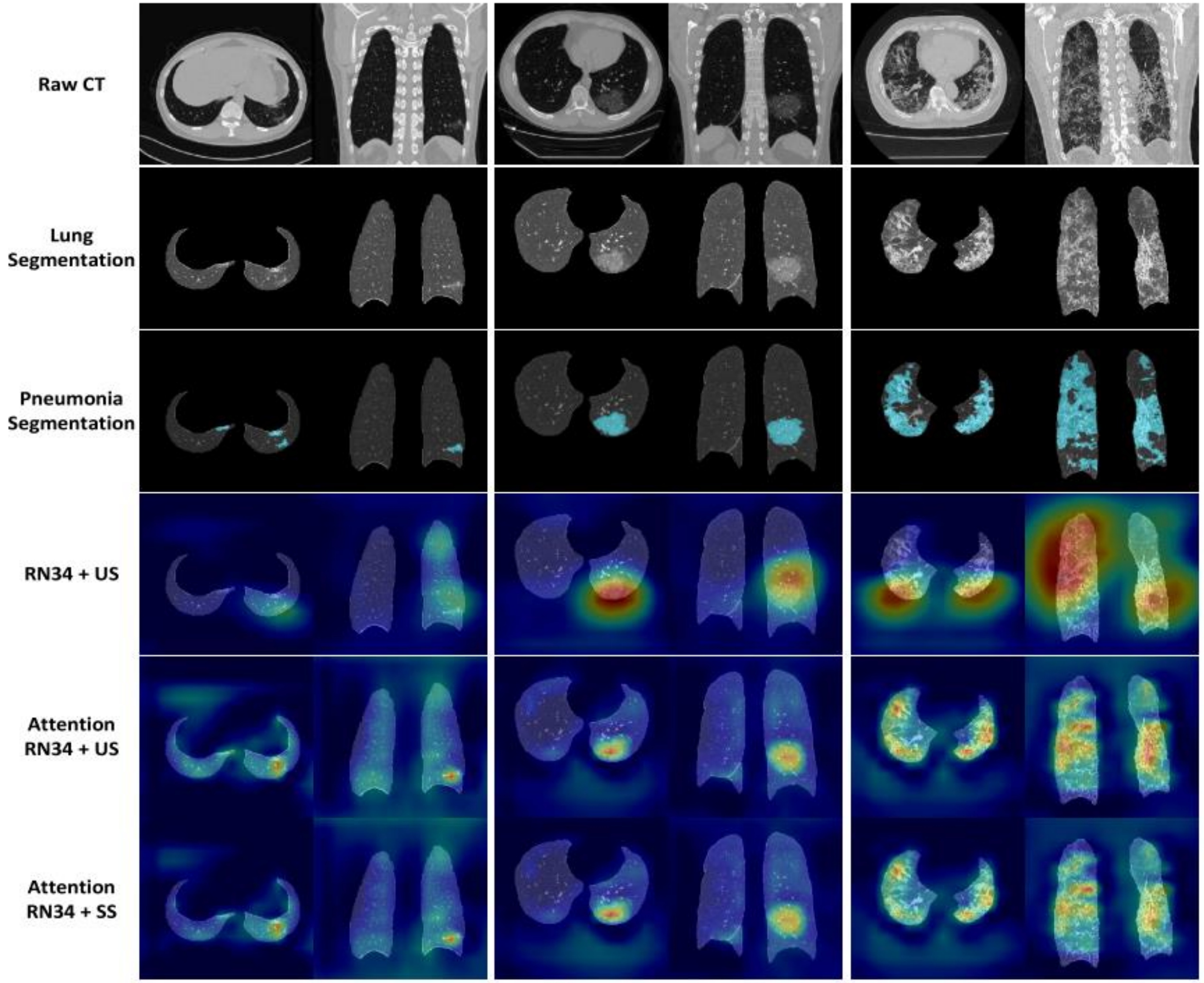


Fig. 5. Visualization results of our methods on three COVID-19 cases from small-infection group ($< \mathbf{0.005}$), median-infection group ($\mathbf{0.005} - \mathbf{0.030}$) and large-infection group ($> \mathbf{0.030}$) of the test set are shown from left to right, respectively. For each case, we show the visualization results in both axial view and coronal view. We show the original images (first row), and the segmentation results of the lung and pneumonia infection regions (2^{nd} and 3^{rd} rows) by the VB-Net tookit [10]. For the attention results, we show the Grad-CAM results of "RN34 +US" (4^{th} row), and the attention maps obtained by our proposed attention module of "Attention RN34 + US" and "Attention RN34 + SS" models (5^{th} and 6^{th} rows).

Cohen, J. P., Dao, L., Morrison, P., Roth, K., Bengio, Y., Shen, B., Abbasi, A., Hoshmand-Kochi, M., Ghassemi, M., Li, H., & Duong, T. Q. (2020). *Predicting COVID-19 Pneumonia Severity on Chest X-ray with Deep Learning*. *8*(December 2019). http://arxiv.org/abs/2005.11856

The extent of lung involvement by ground glass opacity or consolidation for each lung (right lung and left lung separately) was scored as:

0 = no involvement;

1 = <25% involvement;

2 = 25-50% involvement;

3 = 50- 75% involvement;

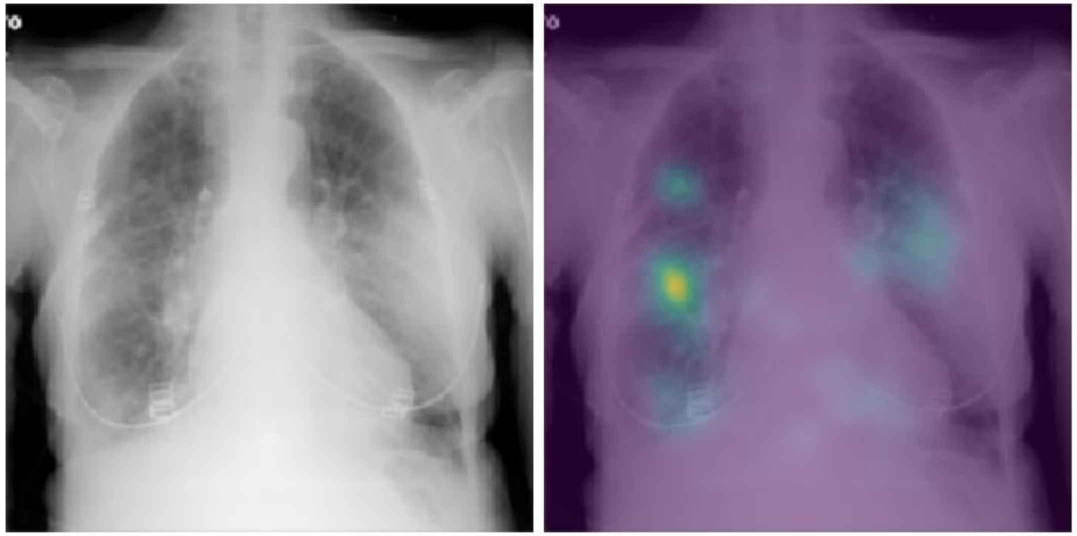
4 = >75% involvement.

The total extent score ranged from 0 to 8 (right lung and left lung together).

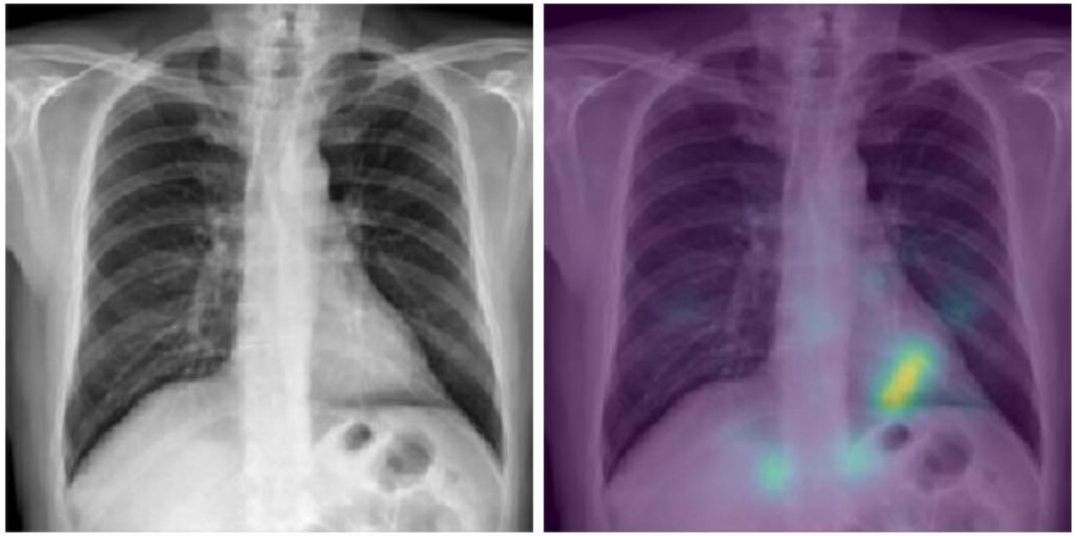
The degree of opacity for each lung (right lung and left lung separately) was scored as:

- 0 = no opacity;
- 1 = ground glass opacity;
- 2 = consolidation;
- 3 = white-out.

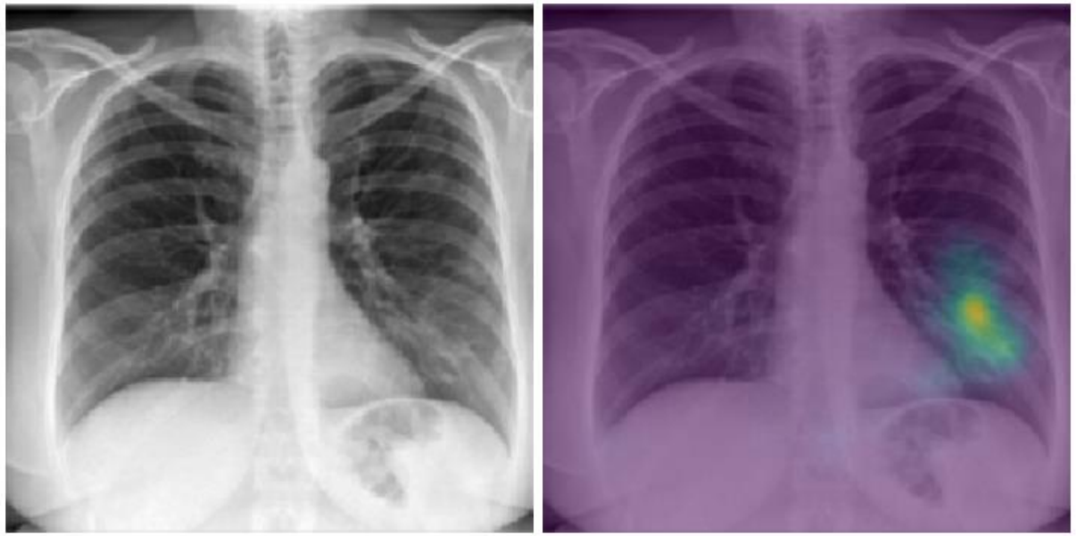
The total opacity score ranged from 0 to 6 (right lung and left lung together).



(a) Geographic Extent Score: 5, Predicted: 5.3



(b) Geographic Extent Score: 0, Predicted: -0.8



(c) Geographic Extent Score: 2, Predicted: 0.62

(d) Geographic Extent Score: 0, Predicted: 1.05

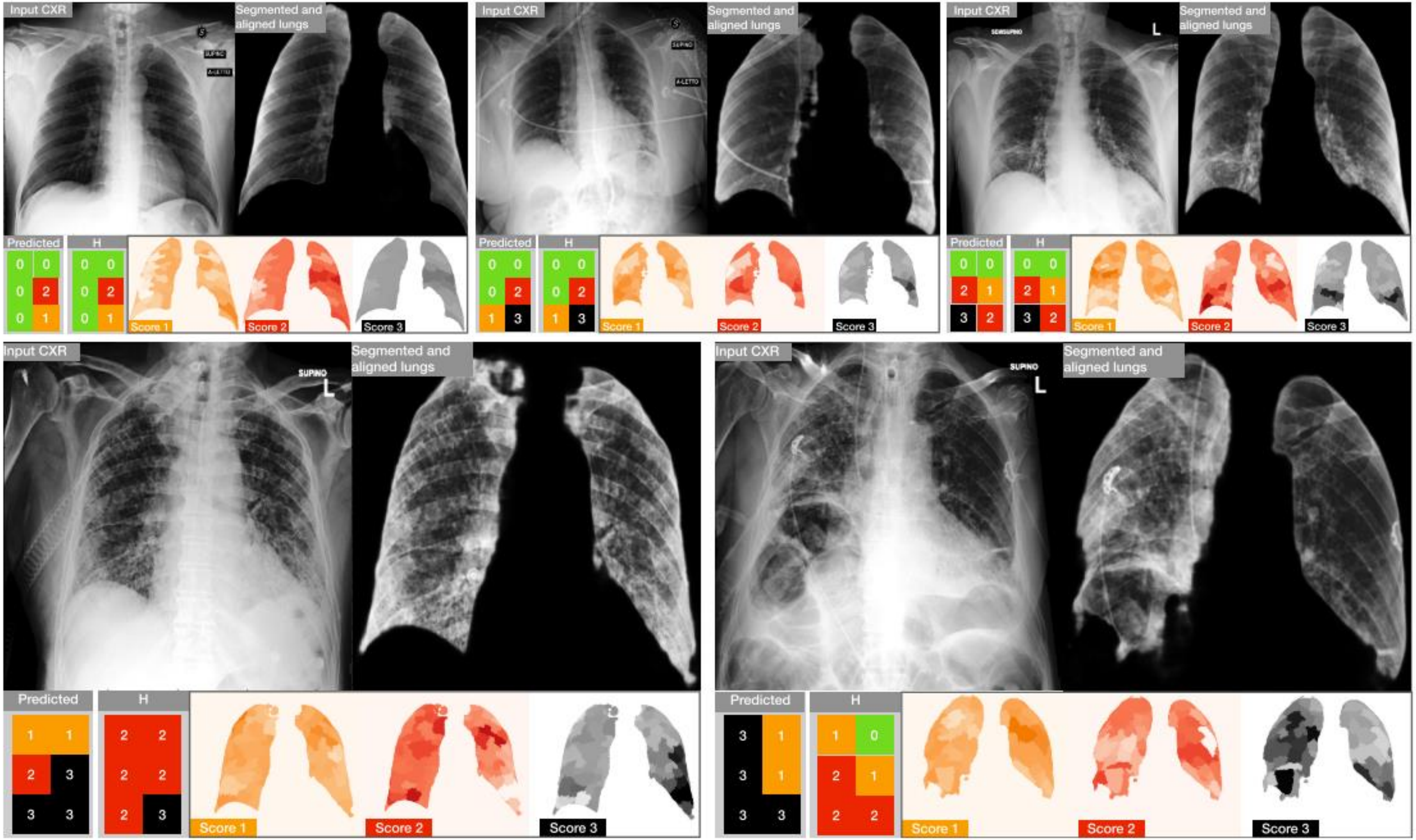
Signoroni, A., Savardi, M., Benini, S., Adami, N., Leonardi, R., Gibellini, P., Vaccher, F., Ravanelli, M., Borghesi, A., Maroldi, R., & Farina, D. (2020). *End-to-end learning for semiquantitative rating of COVID-19 severity on Chest X-rays*. 1–22. http://arxiv.org/abs/2006.04603

According to it, lungs in anteroposterior (AP) or posteroanterior (PA) views, are subdivided into six zones, three for each lung:

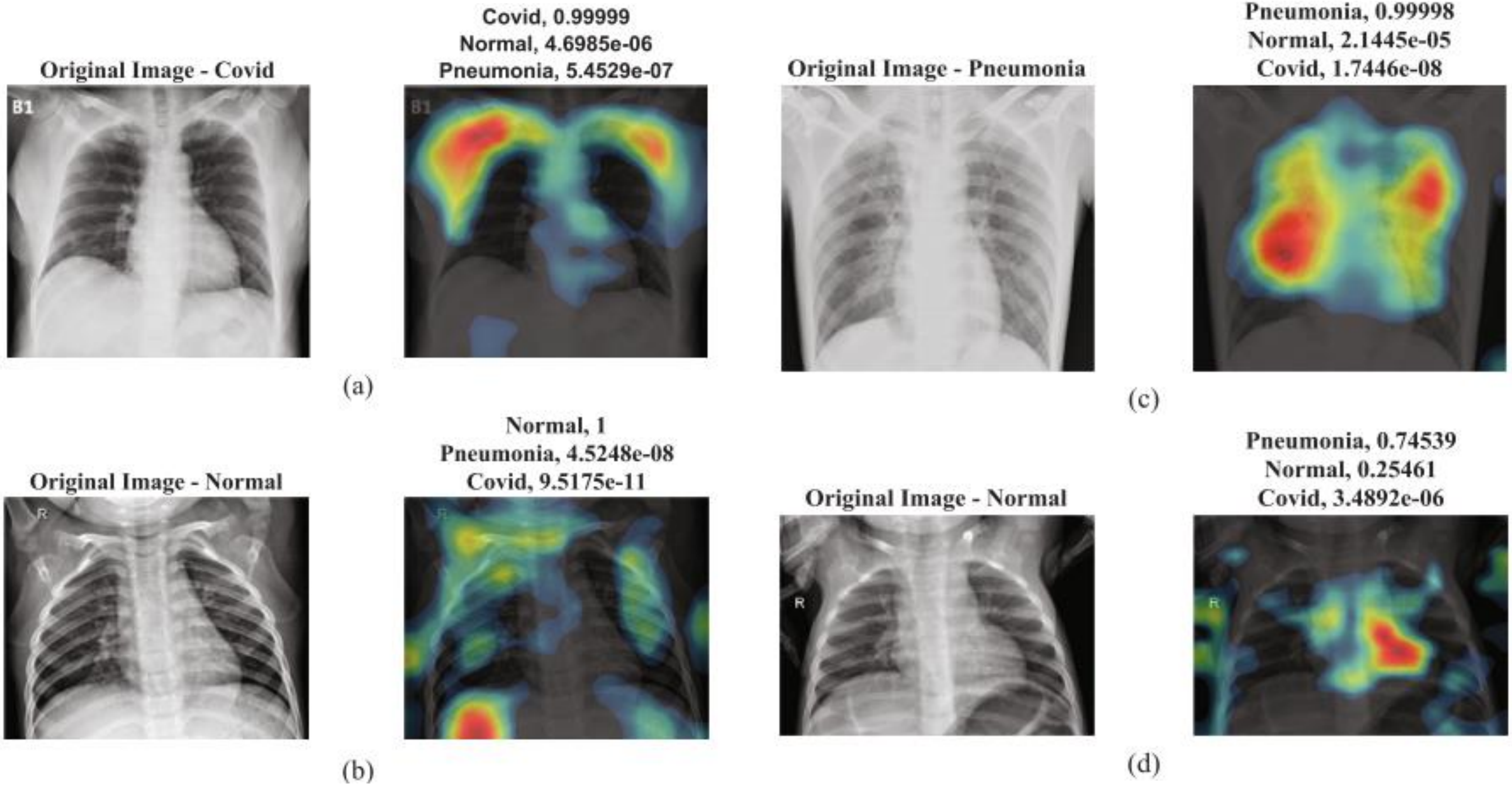
- Upper zones (A and D): above the inferior wall of the aortic arch;
- Middle zones (B and E): below the inferior wall of the aortic arch and above the inferior wall of the right inferior pulmonary vein (i.e., the hilar structures);
- Lower zones (C and F): below the inferior wall of the right inferior pulmonary vein (i.e., the lung bases).

For each zone, a score (ranging from 0 to 3) is assigned, based on the detected lung abnormalities:

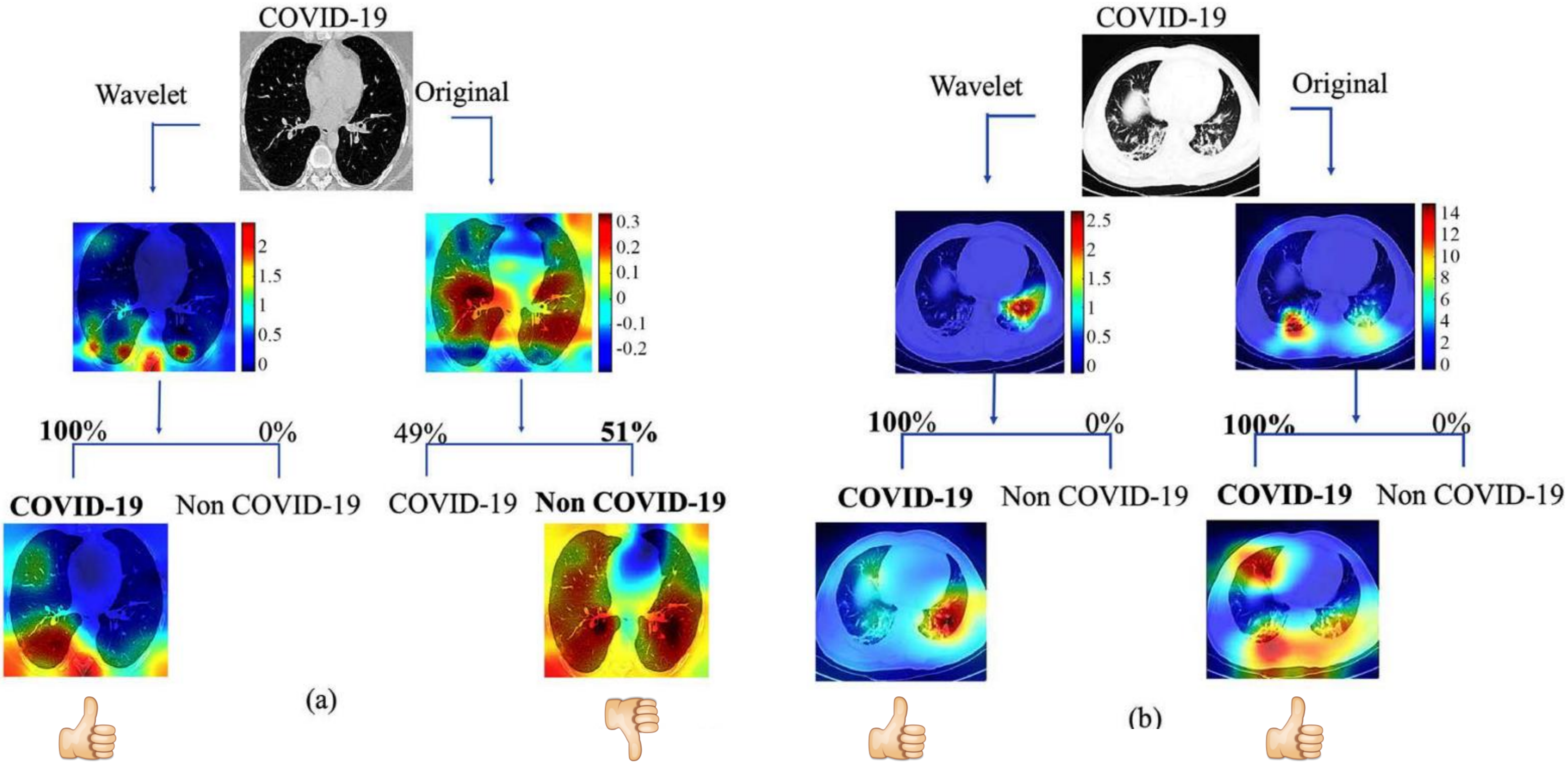
- 0: no lung abnormalities
- 1: interstitial infiltrates
- 2: interstitial (dominant), and alveolar infiltrates
- 3: interstitial, and alveolar (dominant) infiltrate

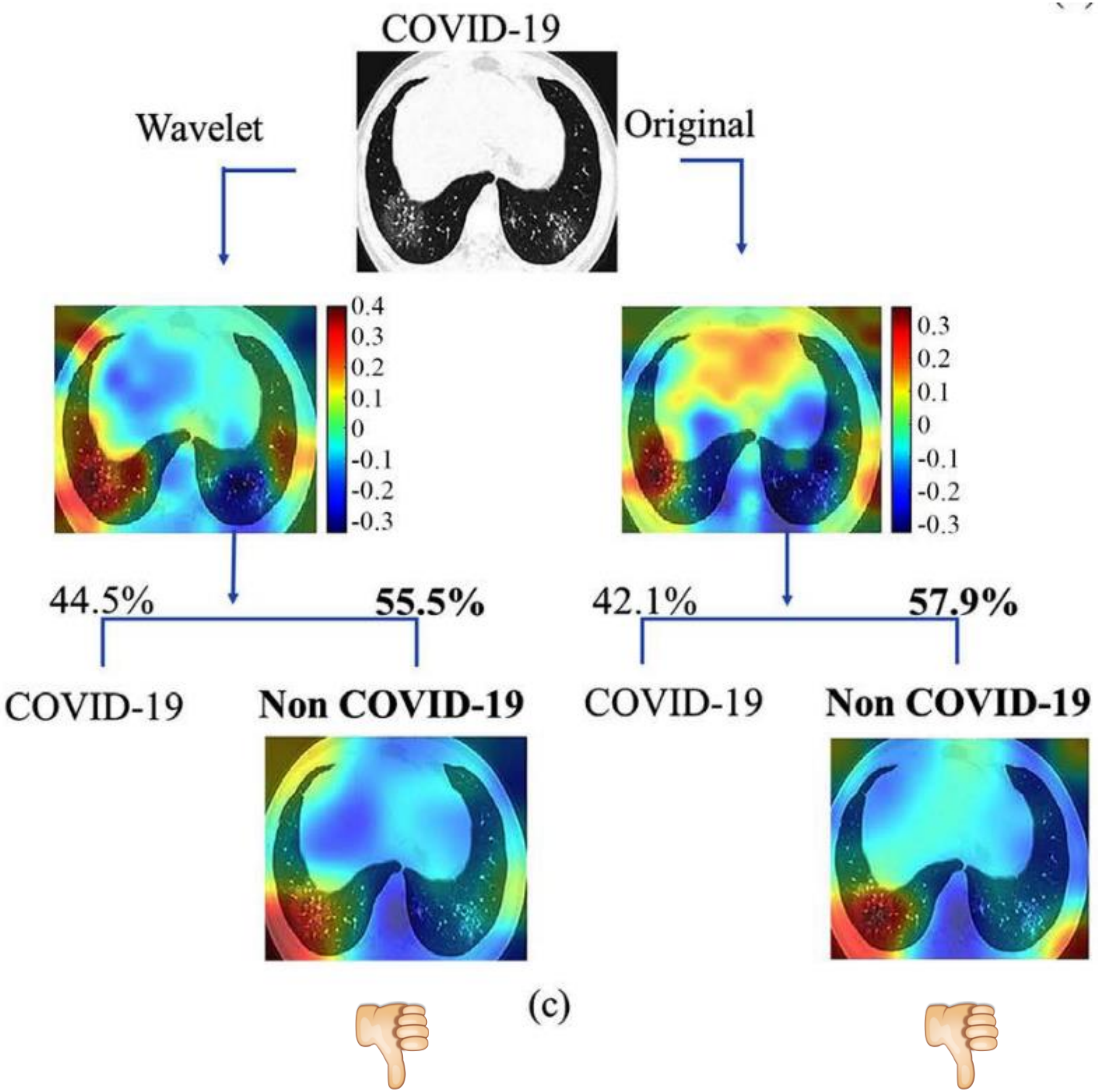

Input CXR
Segmented and aligned lungs
Predicted
0 0
0 2
0 1
H
0 0
0 2
0 1
Score 1
Score 2
Score 3
Input CXR
Segmented and aligned lungs
Predicted
0 0
0 2
1 3
H
0 0
0 2
1 3
Score 1
Score 2
Score 3
Input CXR
L
Segmented and aligned lungs
Predicted
0 0
2 1
3 2
H
0 0
2 1
3 2
Score 1
Score 2
Score 3
Input CXR
SUPINO
L
Segmented and aligned lungs
Predicted
1 1
2 3
3 3
H
2 2
2 2
2 3
Score 1
Score 2
Score 3
Input CXR
SUPINO
L
Segmented and aligned lungs
Predicted
3 1
3 1
3 3
H
1 0
2 1
2 2
Score 1
Score 2
Score 3

Ucar, F., & Korkmaz, D. (2020). COVIDiagnosis-Net: Deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images. *Medical Hypotheses*, *140*(April), 109761. https://doi.org/10.1016/j.mehy.2020.109761



Matsuyama, E. (2020). A Deep Learning Interpretable Model for Novel Coronavirus Disease (COVID-19) Screening with Chest CT Images. *Journal of Biomedical Science and Engineering*, *13*(7), 140–152.




COVID-19
Wavelet
Original
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
44.5%
55.5%
42.1%
57.9%
COVID-19
Non COVID-19
COVID-19
Non COVID-19

(c)

Albert, N. (2020). Evaluation of Contemporary Convolutional Neural Network Architectures for Detecting COVID-19 from Chest Radiographs. *ArXiv.Org*. https://arxiv.org/abs/2007.01108

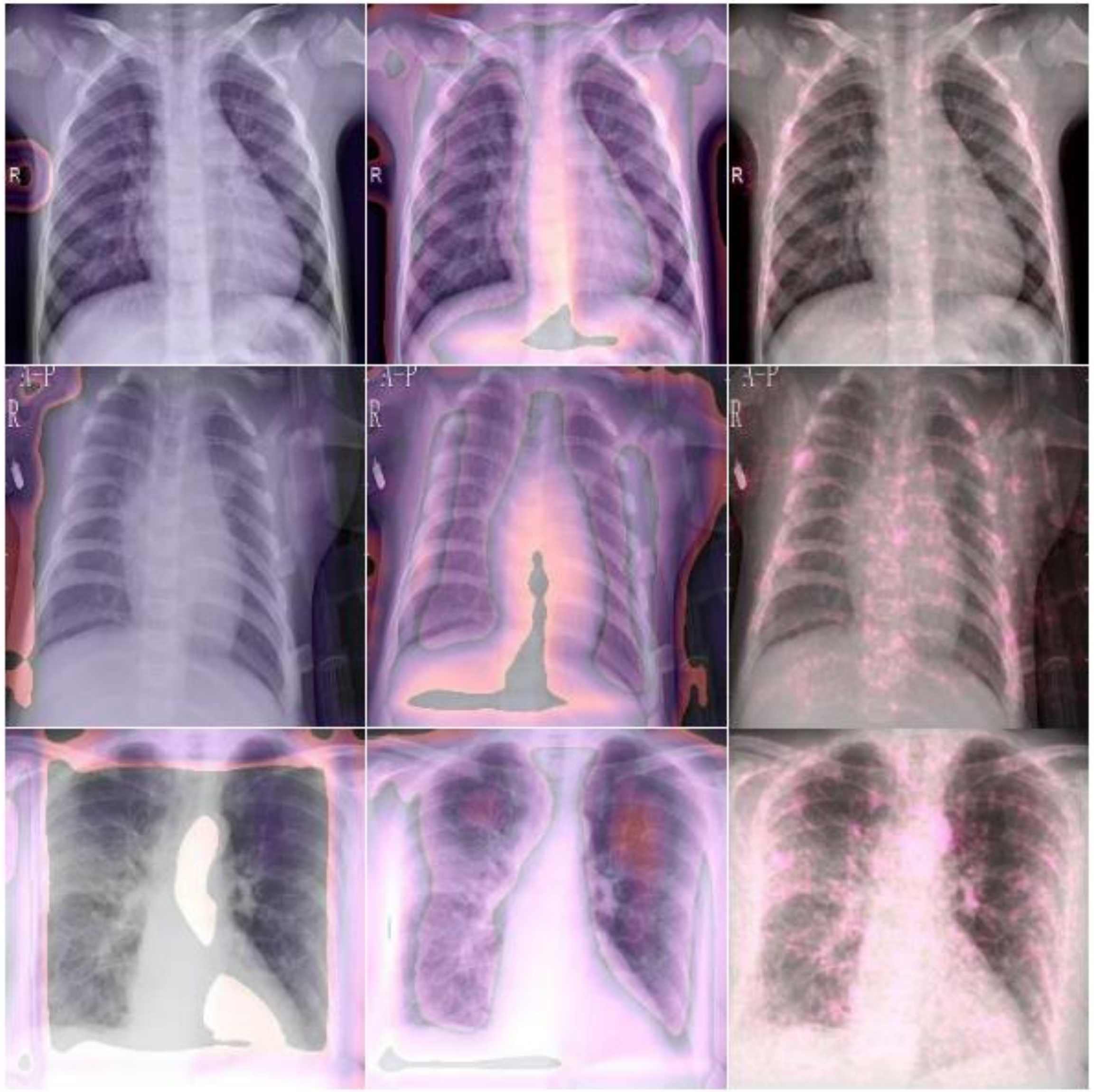


Figure 13. Rows, Top to Bottom: Normal, Bacteria, COVID.
Columns, Left to Right: ImageNet DenseNet, ImageNet ResNet, Efficient model

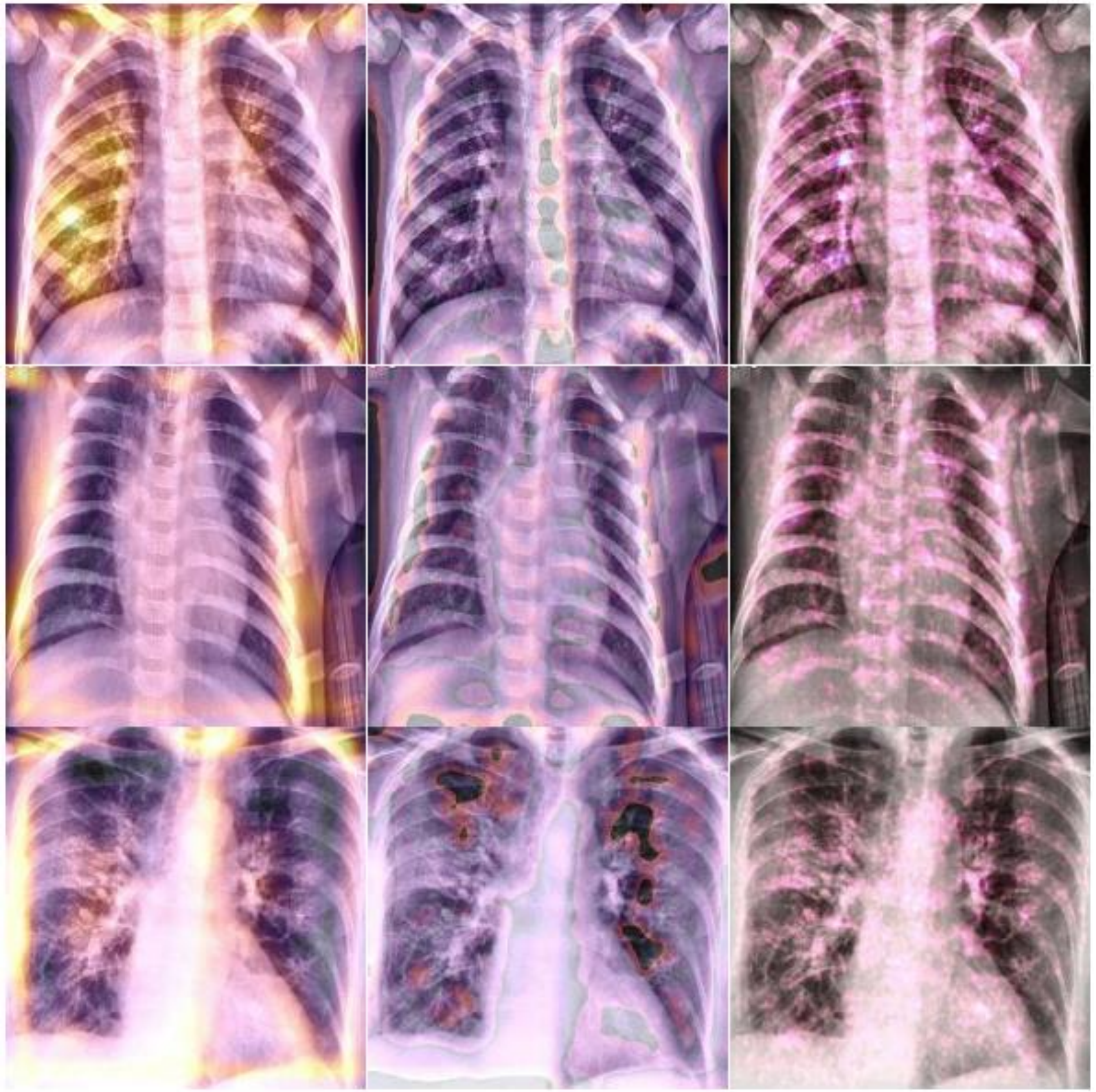
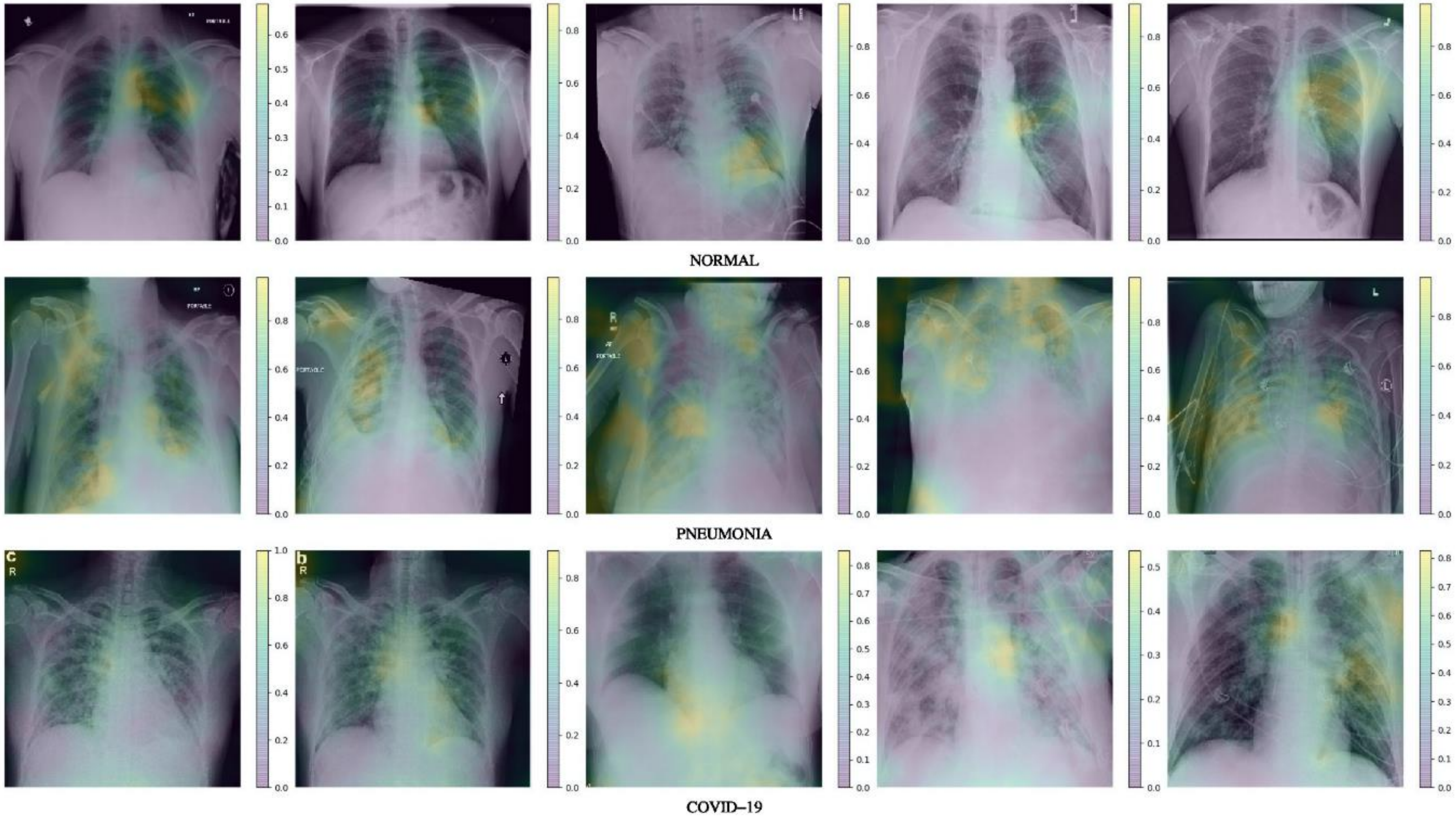


Figure 14. Rows, Top to Bottom: Normal, Bacteria, COVID.
Columns, Left to Right: ImageNet DenseNet, ImageNet ResNet, Efficient model

Khobahi, S., Agarwal, C., & Soltanalian, M. (2020). CoroNet: A Deep Network Architecture for Semi-Supervised Task-Based Identification of COVID-19 from Chest X-ray Images. *MedRxiv*, 2020.04.14.20065722.
https://doi.org/10.1101/2020.04.14.20065722



Wu, Y.-H., Gao, S.-H., Mei, J., Xu, J., Fan, D.-P., Zhao, C.-W., & Cheng, M.-M. (2020). *JCS: An Explainable COVID-19 Diagnosis System by Joint Classification and Segmentation*. 1–11. http://arxiv.org/abs/2004.07054

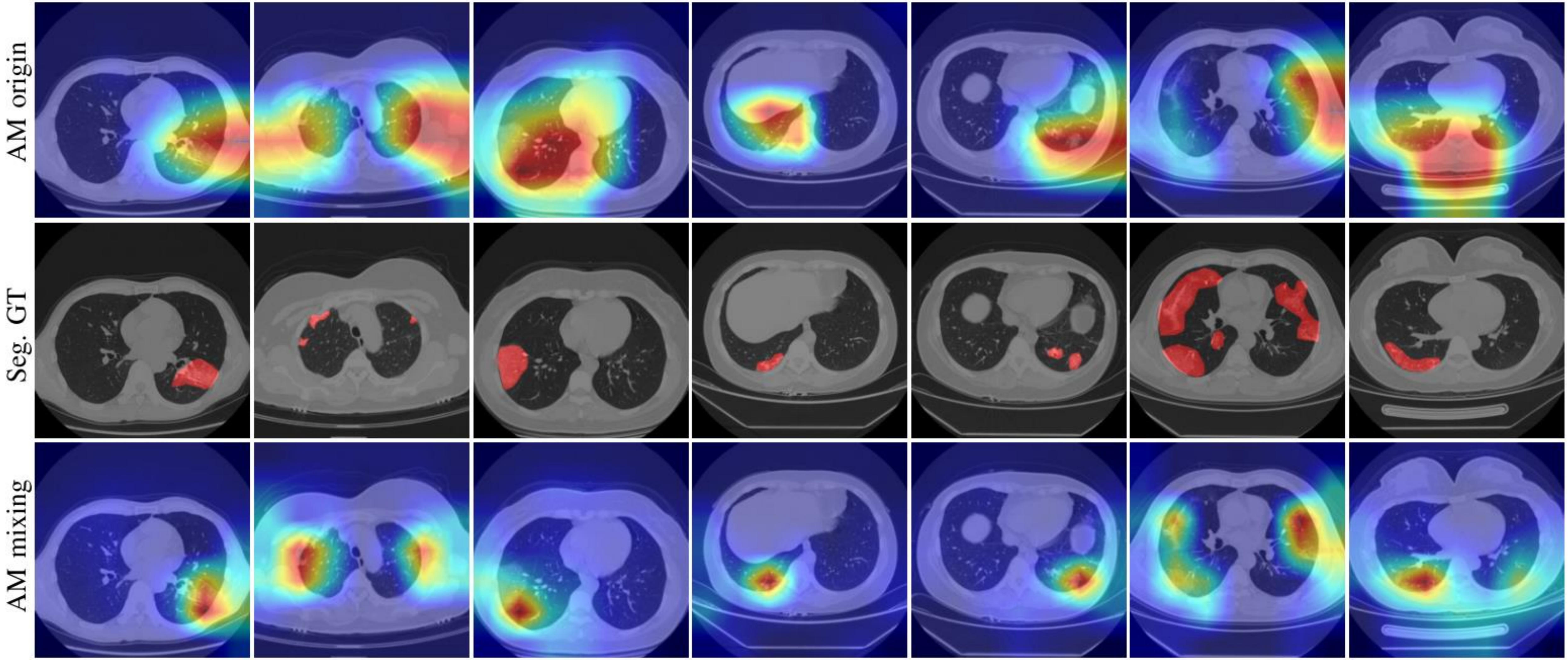


Figure 8. **Visualizations of activation mapping (AM)**. AM origin (mixing) means the AM of models trained without (with) image mixing technique [49].

Ahsan, M. M., Gupta, K. D., Islam, M. M., Sen, S., Rahman, M. L., & Hossain, M. S. (2020). *Study of Different Deep Learning Approach with Explainable AI for Screening Patients with COVID-19 Symptoms: Using CT Scan and Chest X-ray Image Dataset*. http://arxiv.org/abs/2007.12525

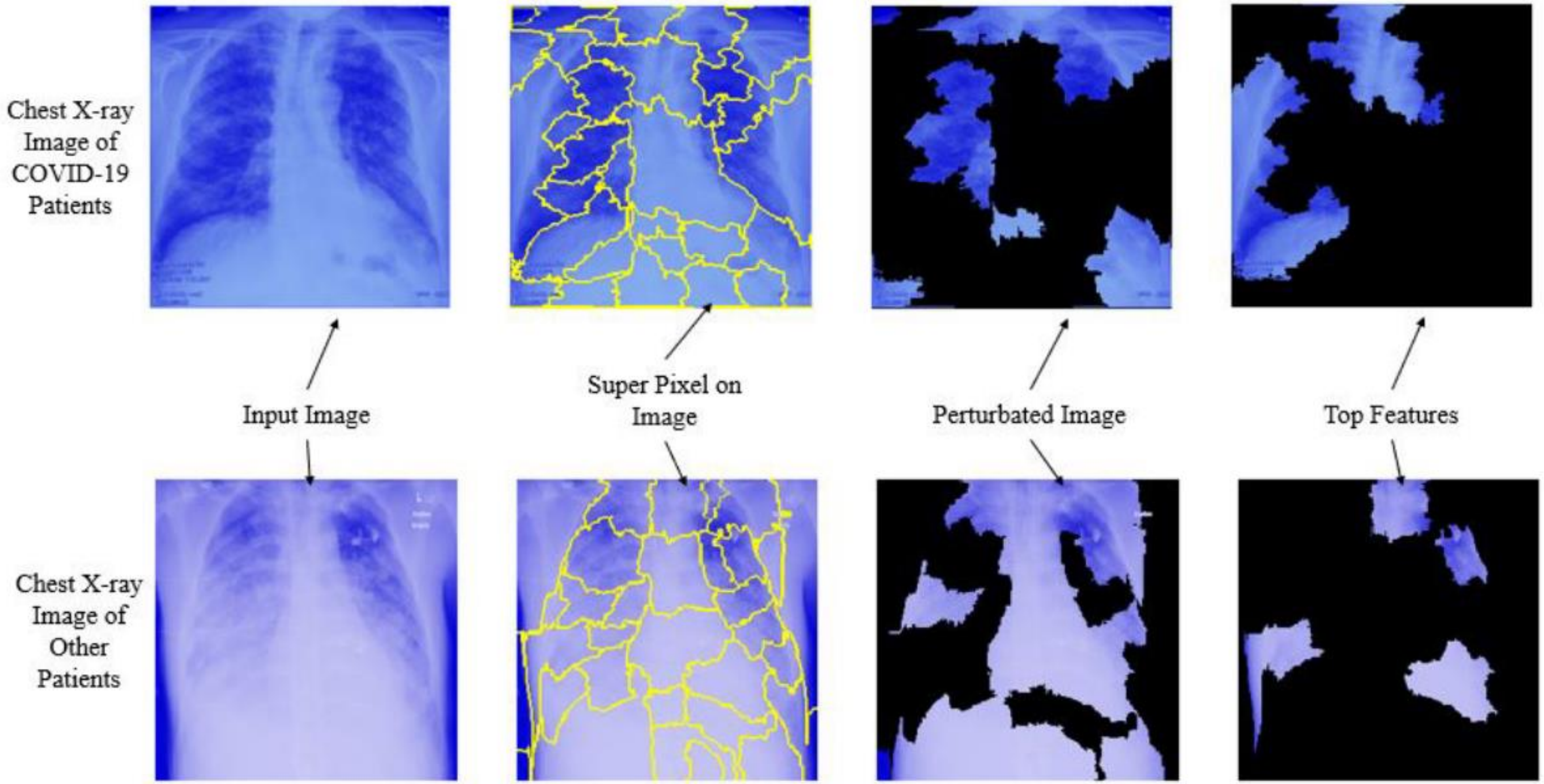


Figure 15: Overall prediction analysis using LIME

Jaiswal, A. K., Tiwari, P., Rathi, V. K., Qian, J., Pandey, H. M., & Albuquerque, V. H. C. (2020). COVIDPEN: A Novel COVID-19 Detection Model using Chest X-Rays and CT Scans. *MedRxiv*, 2020.07.08.20149161. https://doi.org/10.1101/2020.07.08.20149161

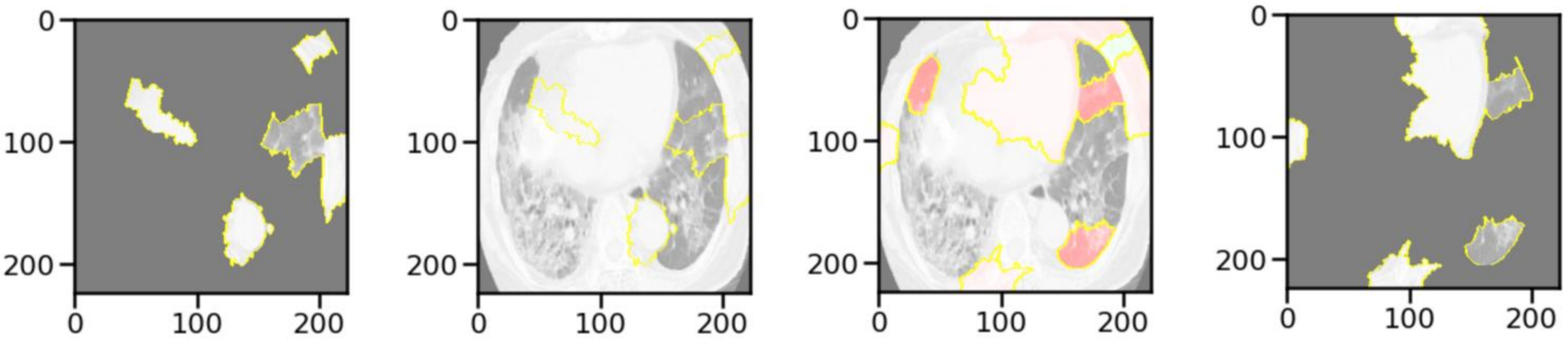


Fig. 6: **Model Interpretability - COVID-19 CT Scan Dataset**

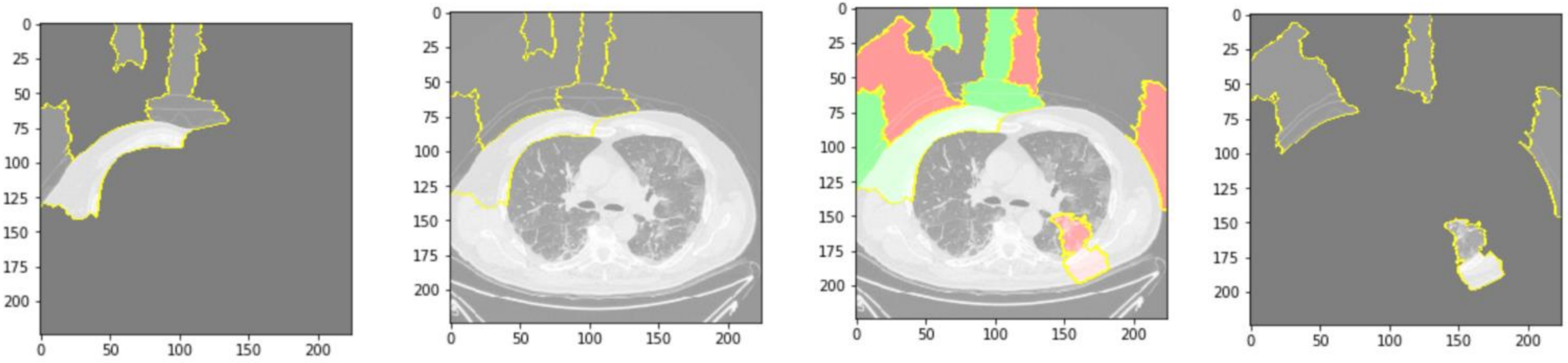


Fig. 7: **Model Interpretability - COVID-19 Chest Radiographs Dataset**

The regions shaded in pink and green detect superpixels that contributed to and against prediction of chest radiographs and CT scans.

Wang, L., & Wong, A. (2020). *COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images*. 1–12. http://arxiv.org/abs/2003.09871

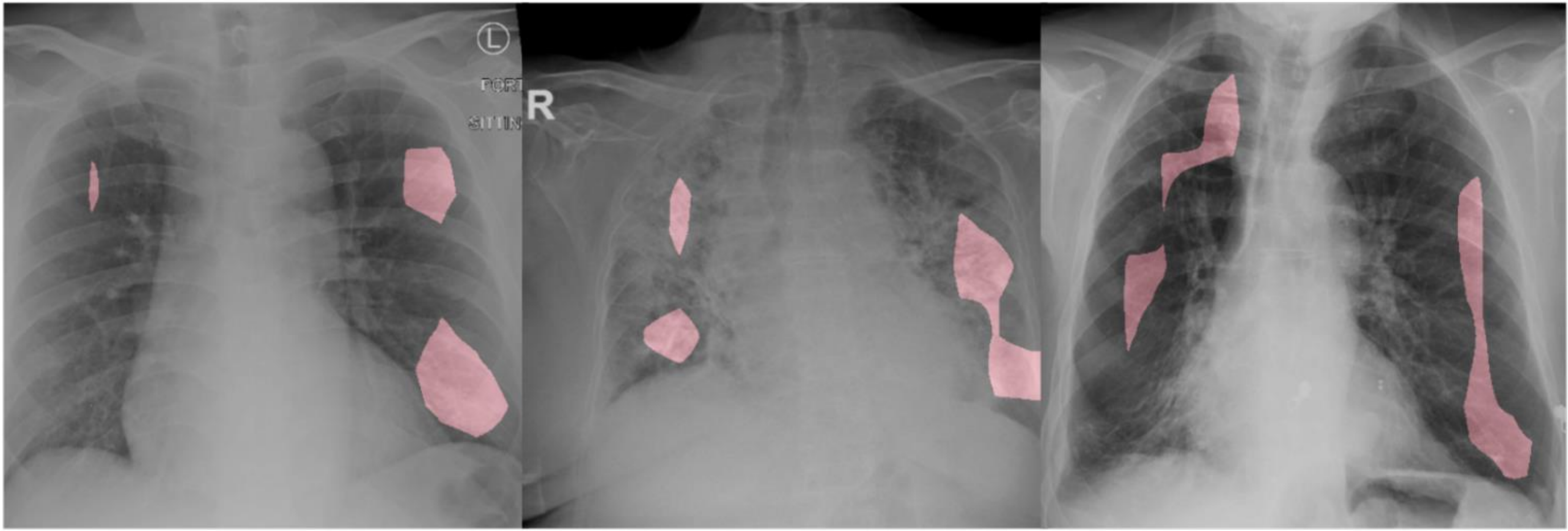


Figure 7. Example CXR images of COVID-19 cases from several different patients and their associated critical factors (highlighted in red) as identified by GSInquire[37].

Zokaeinikoo, M., Mitra, P., Kumara, S., & Kazemian, P. (2020). AIDCOV: An Interpretable Artificial Intelligence Model for Detection of COVID-19 from Chest Radiography Images. *MedRxiv*, 2020.05.24.20111922. https://doi.org/10.1101/2020.05.24.20111922

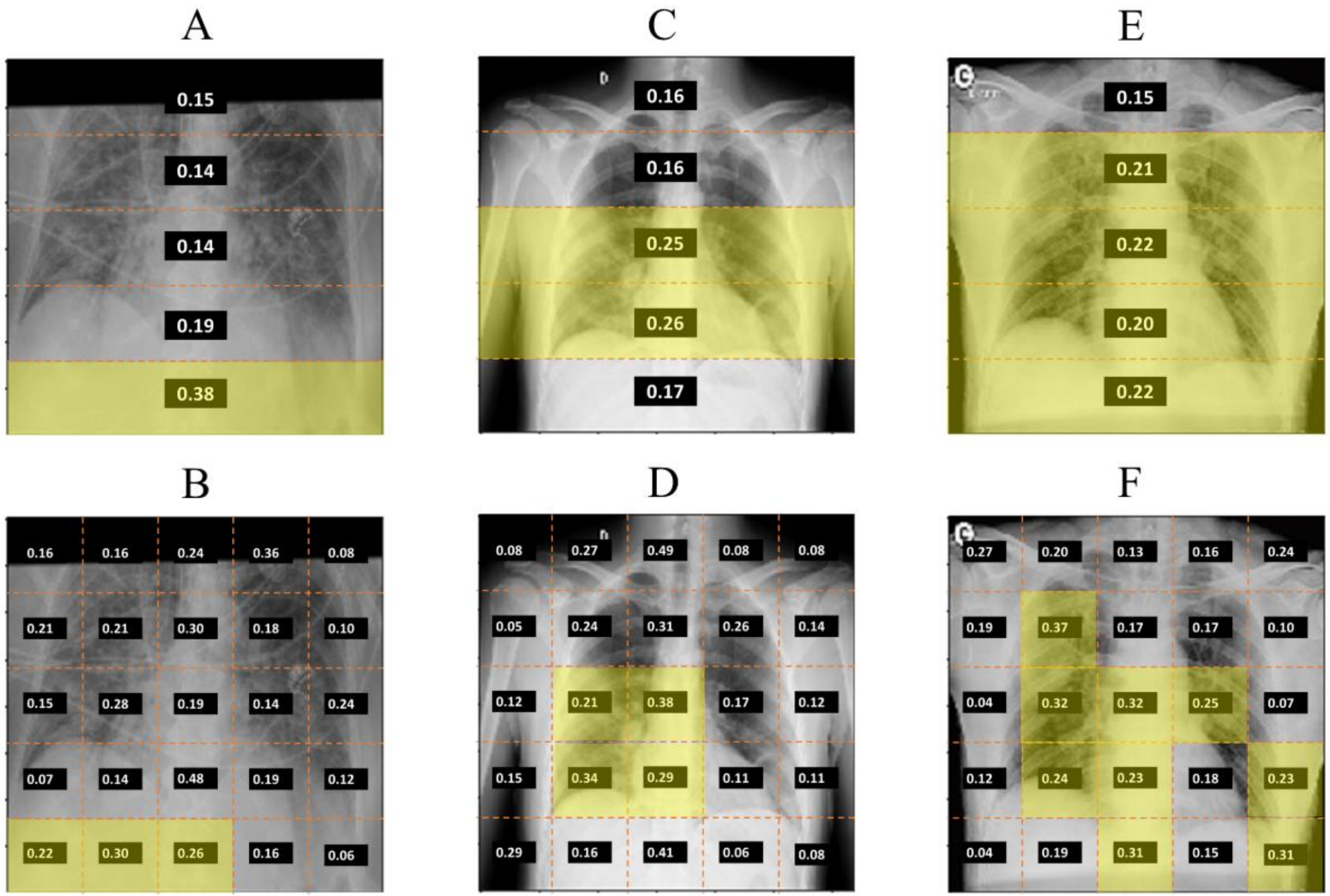


Figure 6: Attention scores for different zones of the lung (horizontal level) and different blocks of the image for 3 patients with COVID-19. Signs of COVID-19 were detected in the lower zone for Patient 1 (A-B), middle zone for Patient 2 (C-D), and lower and middle zones for Patient 3 (E-F).