# AI EXPLAINABILITY IN AUTOMATED GROUND GLASS OPACITY (GGO) SEGMENTATION USING CHEST CT SCANS


*Shereen Fouad* [⋆] *Paula Atim* [⋆] *Arvind Rajasekaran* [†] *Pankaj Nagori* [†] *John Morlese* [†] *Bahadar Bhatia* [†]

[⋆] School of Computer Science and Digital Technologies, Aston University, UK
[†] Sandwell and West Birmingham Hospitals NHS Trust, UK


Deep learning (DL) technology has been recently considered as an effective tool for ground-glass opacity (GGO) segmentation from chest computed tomography (CT) image. The segmentation results aid the early detection and monitoring of major conditions, including COVID-19, acute chest malignancies, and other lung conditions. Our previous work in [1], introduced an improved U-Net model with a pretrained ResNet-50 for GGOs segmentation in CT images. The dataset consists of 62 pseudonymized chest CT studies with a slice thickness of 2 mm from Covid-19 patients, annotated by radiologists, utilising the BSTI guidelines. Results revealed that the proposed ResNet-50U-Net outperformed classic U-Net and DenseNet-121U-Net networks in identifying the GGO regions, achieving Dice similarity, Precision, and Recall scores of 0.71, 0.63, and 0.83, respectively. To improve transparency and confidence, we applied two widely used visual Explainable AI (XAI) techniques (Grad-CAM++ (Gradient-weighted Class Activation Mapping) and LIME (Local Interpretable Model-agnostic Explanations)), which we previously evaluated by 26 medical doctors in a user study in [2]. This work extends our previous two studies by comparing the Grad-CAM++ and LIME in explaining GGOs detection in CT scans. During experiments, both XAI methods were run repeatedly on the same images. Results are shown in Fig1, Grad-CAM++ highlights (using heatmaps) the areas that contribute most towards the model prediction using the gradients of the CNN's last layer. LIME explains the AI output by perturbing parts (superpixels) of the input image and analyzing their impact on the output. It identifies influential regions (highlighted in red in image Fig 1) to the DL-based model.

Overall, XAI results obtained from LIME and Grad-CAM are visually inconsistent. Our results show that GradCAM++ produces heatmaps that highlight regions of high significance in a continuous and spatially coherent manner, making it easier to interpret the GGO segmentation task. LIME offers simpler, model-agnostic explanations by segmenting input into interpretable superpixels (shown in red). However, it consistently activates irelevant regions (red superpixels) and fails to detect other more relevant regions, leading to less comprehensive or less accurate interpretability compared to Grad-CAM++. These results align with our previous findings in [2], where medical doctors demonstrated a clear preference for Grad-CAM over LIME in chest radiology imaging tasks. Specifically, 80% of users favored Grad-CAM due to its ability to provide more detailed, coherent, and consistent insights, which enhanced both trust and interpretability in the model's predictions.

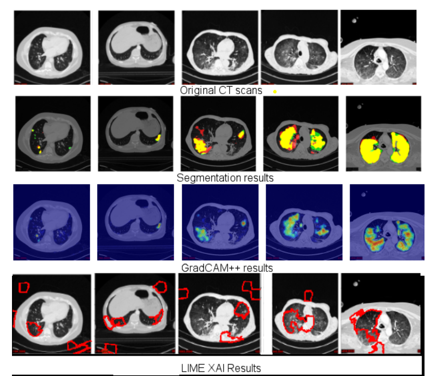


**Fig. 1**. (First row) original CT scans (second row) segmentation results on five CT examples. Green represents the ground truth masks, red represents the automatic segmentation, and yellow is the overlap between the automatic and manual segmentation. Third and fourth rows show GradCAM++ and LIME XAI results on ResNet-50U-Net model, respecively

## 1. REFERENCES


[1] Paula Atim, Shereen Fouad, Sinling Tiffany Yu, Antonio Fratini, Arvind Rajasekaran, Pankaj Nagori, John Morlese, and Bahadar Bhatia, "Explainable deep learning framework for ground glass opacity (ggo) segmentation from chest ct scans," in *Medical Imaging and Computer-Aided Diagnosis*. Springer, 2024.

[2] Izegbua E. Ihongbe, Shereen Fouad, Taha F. Mahmoud, Arvind Rajasekaran, and Bahadar Bhatia, "Evaluating explainable artificial intelligence (xai) techniques in chest radiology imaging through a human-centered lens," *Plos one*, vol. 19, no. 10, pp. e0308758, 2024.