# Explainable Deep Learning Framework for ground glass opacity (GGO) Segmentation from Chest CT scans

Paula Atim †[1], Shereen Fouad † *[1], Sinling Tiffany Yu[1], Antonio Fratini[1], Arvind Rajasekaran[2], Pankaj Nagori[2] John Morlese[2], and Bahadar Bhatia[2,3]
† These authors contributed equally to this work. *s.fouad@aston.ac.uk

[1] College of Engineering and Physical Sciences, Aston University, UK
[2] Sandwell and West Birmingham Hospitals NHS Trust, UK,
[3] University of Leicester, UK

**Abstract.** Segmenting ground glass opacities (GGO) from chest computed tomography (CT) scans is crucial for early detection and monitoring of lung diseases. This includes lung infections and acute alveolar malignancies. However, GGO segmentation is a challenging task in chest radiology as GGOs often exhibit a range of characteristics and displays low-intensity contrast with adjacent structures in CT images. This study introduces a novel deep learning framework for segmenting GGOs in CT scans using ResNet-50U-Net, which is an improved U-Net model with a pretrained ResNet-50 to enhance feature extraction. A total 62 CT pseudoanonymised images were collected from patients with Covid-19, annotated by experienced radiologist, and further processed for analysis. Our experimental results demonstrate that the proposed ResNet-50U-Net outperforms the standard U-Net as well as DenseNet-121U-Net architectures in detecting the GGO locations with Dice similarity score, Precision, and Recall of 0.71, 0.63, and 0.83, respectively. Unlike current deep learning-enabled methods for GGO segmentation, which face trust challenges due to their "black-box" nature, our approach integrates a post-hoc visual explainability feature through the GradCAM++ (Gradient-weighted Class Activation Mapping) algorithm. This tool highlights significant regions within the Chest CT scans that impacts the model's decision, providing beneficial insights into the segmentation process.

**Keywords:** GGO segmentation · Computed Tomography (CT) · Deep Learning · U-Net · Explainable Artificial Intelligence

## 1 Introduction and Background

Ground-glass opacity (GGO), as defined by the Fleischner Society, refers to a hazy increase in lung density on high-resolution Computed Tomography (CT) that does not obscure the underlying vessels or bronchial walls [7]. If the vessels are obscured, the term "consolidation" is used. GGO are often associated

with several lung disorders ranging from benign conditions such as inflammation and fibrosis through preinvasive lesions like atypical adenomatous hyperplasia or adenocarcinoma in situ which are early-stage tumors. Furthermore, in COVID-19 detection, GGO is a key marker of viral pneumonia and lung damage, often appearing early in the disease. Accurate identification of GGOs in CT images is crucial for the early detection and effective management of COVID-19 and other lung diseases. It enhances the chances of timely intervention, improving patient survival rates while reducing the need for more invasive treatment methods. However, detecting GGOs boundaries in CT scans may prove difficult owing to their blurred appearance and diversity in GGO shapes along with a lack of contrast when compared to neighboring structures [9]. Additionally, segmenting the GGOs in CT images manually is a time-intensive process and susceptible to both intra- and inter-observer variability.

The automatic segmentation (detection) of GGOs in CT images have been investigated widely in the literature. Earlier studies have utilised a combination of traditional image segmentation methods, including image processing, statistical, and classic machine learning algorithms. For instance, [3] uses Gaussian mixture model and high-pass filter to extract GGO regions in CT images. False positives of this method were then reduced by applying shape filtering and local feature analysis. [13] applies a series of image processing techniques, including lung segmentation using an edge-searching method, nodule enhancement through image accumulation, and ROI detection using support vector machine (SVM) algorithm. Another research in [1] uses morphological reconstruction technique for GGOs segmentation in CT scans, followed by statistical analysis and a joint feature technique, which reduces overlap in grey-scale values. However, those methods suffered from low segmentation accuracy, long processing time and poor reliability. More recently, Deep Learning (DL) has shown remarkable performance in segmenting GGOs from CT images. These tools can learn from labeled training data to make diagnostic inferences on new data inputs. Typically, experts such as radiologists manually label the training data to mark the exact locations of GGO regions. A study in [10] uses Deep Convolutional Neural Network (DCNN) method to extract GGO in chest CT images. Key steps include image preprocessing, lung area extraction, vessel and bronchial removal, and candidate region segmentation. The DCNN is then applied to enhanced images, achieving an 86.05% true positive rate on the Lung Image Database Consortium (LIDC). The study in [18] proposes a DL system for early identification of GGOs using a combination of AlexNet, GoogLeNet, and Resnet-50 networks on Hospital CT images. The research proposes a novel preprocessing technique involving RGB superposition to enhance differentiation between nodules and normal tissues, achieving an 88% accuracy in lung nodule detection. [5] introduces COVID-19 lung CT infection DL-enabled segmentation network, called Inf-Net, which enhances the detection of GGOs through implicit reverse attention and explicit edge-attention mechanisms. Despite the success of DL methods in medical imaging segmentation tasks, these models are often criticized for their "black-box" nature, which makes it difficult to interpret how they arrive at specific decisions.

U-Net is a convolutional neural network (CNN) architecture designed for medical image segmentation tasks [15]. U-Net uses an encoder-decoder structure with skip connections to capture both contextual and spatial information for accurate pixel-wise predictions. In recent years, several studies have combined U-Net with pretrained architectures in the encoder part to enhance image segmentation. For example, in [14], U-Net was combined with three different backbones: VGG16, ResNet50, and Xception, and applied to brain tumor segmentation using MRI images. Experiments revealed that the combination of ResNet50 and U-Net, referred to as ResNet-50U-Net, achieved the best segmentation results. However, the segmentation performance of the ResNet-50U-Net has not been previously studied in GGO segmentation tasks applied to CT images.

This study aims to explore the performance of ResNet-50U-Net in the context of GGO segmentation and compare the results against other U-Net-based pretrained architectures as well as standard U-Net models. To improve segmentation performance, we modified the ResNet-50U-Net architecture by integrating residual connections in its decoder blocks. This integration helps preserve spatial information and improve the quality of GGO segmentation in CT scans. Furthermore, we augment our segmentation model with an Explainable AI (XAI) algorithm, specifically the Grad-CAM++ (Gradient-weighted Class Activation Mapping) algorithm [4], which highlights the image regions that influence the model's decision through heat maps, aiming to balance explainability with strong GGO segmentation performance. The application of XAI in GGO segmentation in CT scans remains unexplored, which is crucial in clinical environments, where trust and interpretability are critical [12,6,2].

## 2 Data collection and preprocessing

In this study, the dataset was gathered in 2020 from hospitals within the Sandwell and West Birmingham National Health System Trust in the United Kingdom. The dataset includes 62 pseudonymized chest CT studies with 2 mm slice thickness. Studies were reported through standard-of-care where the clinical referrer suspected the presence Covid-19. The study design was cross-sectional survey, double-blind, retrospective, non-interventional and at a single NHS Trust. The GGOs were identified and labelled (using ITK-SNAP v3.8.0.) by expert reviewers (>twenty years' experience) including a general radiologist and two chest speciality radiologists, utilising the British Society of Thoracic Imaging (BSTI guidelines, version 2 13.04.2020) (Fig 1.(a) and (b)). Studies were pseudonymised on-premise using open-source tools: the Radiological Society of North America Clinical Trials Processor. The study received ethical approval from the Hospital as well as Aston University's ethics committee. All procedures were carried out in accordance with applicable guidelines and regulations.

The GGOs binary regions was performed using Materialise Mimics 26.0 (Materialise, Leuven, Belgium). The technique isolates the GGOs based on their specific Gray values, which represents the tissue density in Hounsfield units (HU)

for CT images. The dynamic region grow tool was used to select an initial seed point where the GGO is located and neighbouring pixels within a set minimum and maximum deviation were included in the segmentation (100-150 HU). The multiple slice edit tool was then used to refine the mask by adjusting over and under-segmented regions. The generated binary regions (Fig 1.(c)) were compared against the ground truth images obtained by radiologist (Fig 1.(b)) to ensure accuracy. Final masks were saved in BMP format and converted into binary format using Python, where Black (0) represented the background and White (255) represents the GGO and foreground.

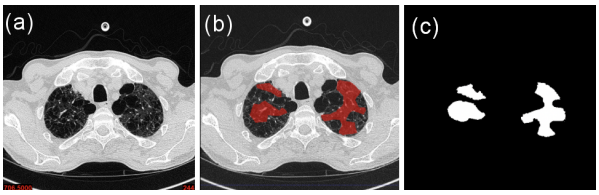


**Fig. 1.** (a) original CT scan, (b) manual annotation of GGO marked in red by a radiologist and (c) GGO mask generated using Materialise Mimics 26.0.

Given our relatively small dataset and the substantial data requirements of DL models for effective training, we employed data augmentation as a key preprocessing technique. This approach artificially enhances the diversity of the training dataset by applying various transformations to the input data. To achieve this, we used ImageDataGenerator from TensorFlow which is used to increase the size and enhance the training dataset. The augmentations include rotation (up to 15 degrees), width and height shifts (10%), shear transformations, zoom (80% to 120%), and horizontal flips, with the fill mode set to 'reflect' to maintain image continuity. Masks were augmented similarly, with an added preprocessing step converting them to binary format and they were paired to their corresponding image using custom generators.

## 3 The Segmentation Model

This study proposes a DL-based approach for GGO segmentation in CT scans. The proposed model leverages the strengths of both baseline U-Net and Resnet-50 architectures for segmentation tasks. The performance of the model will be compared against standard U-Net and DenseNet121U-Net in the context of GGO segmentation tasks.

### 3.1 U-Net Architecture

Our U-Net model takes a 512x512 input image, encoding it through two convolutional layers with ReLU activations, followed by max pooling to reduce dimensions and capture multi-scale features. The image size shrinks to 32x32 with

512 channels. These features are passed through a Double Convolution block in the bottleneck, producing a 32x32 map with 1024 channels. In the decoder, transposed convolutions upsample the feature maps back to 512x512, using skip connections to concatenate with encoder features. A final Conv2d layer refines the output, with ReLU activations driving pattern learning throughout.

### 3.2 DenseNet-121U-Net Architecture

DenseNet DL model [11] uses dense blocks and has a feed-forward structure. In this study, we explore DenseNet121 and integrate it into a U-Net framework, forming a DenseNet-121U-Net. In this model, the encoder uses a pretrained DenseNet-121, which processes a 512x512 input image through multiple layers, extracting increasingly abstract features at each stage with 64, 256, 512, and 1024 channels. ReLU activation functions are applied after each convolution to introduce non-linearity. The decoder then mirrors this process, upsampling features from to 1024 channels back to 64 channels in stages. At each step, it uses transposed convolutions and skip connections from the encoder to retain spatial details, and additionally, it employs dense residual additions to further enhance feature representation and maintain important information throughout the upsampling process. Finally, a 1x1 convolution produces the segmentation map, which is upsampled to match the original input size, making this model effective for precise GGO segmentation tasks.

### 3.3 The proposed ResNet-50U-Net Architecture

ResNets [8] is a popular CNNs-based classification model that leverages residual blocks, which enable the network to learn residual functions that map inputs to the desired outputs to solve challenging image recognition problems. ResNet-50 [17] is a 50 layers CNNs architecture pre-trained on the ImageNet data set.

The proposed ResNet-50U-Net model combines the U-Net model with Resnet-50 [17] as a backbone to obtain the best GGO detection. The Resnet-50-U-Net segmentation architecture (see Fig 2) comprises an encoder to extract features from the input image by downsampling, using the ResNet50 backbone. The encoder reduces spatial dimensions while increasing the number of filters from 64 to 1024, capturing complex features at multiple scales. The decoder upsamples the feature maps using transposed convolutions and incorporates skip connections, which concatenate features from the encoder to refine the reconstruction. It gradually restores the original image resolution, enabling precise pixel-wise segmentation. One unique aspect of the proposed ResNet-50-U-Net model is that it integrates residual connections in its decoder blocks. These residual connections add the upsampled input back to the output of each convolutional block, which consists of two convolution layers with 3x3 filters, batch normalization, and ReLU activation. This integration helps preserve spatial information and improve gradient flow, which is crucial for maintaining stability during training in deep networks. The added residual connections allow the model to bypass some layers, preventing vanishing gradients and helping the network retain

more useful low-level features, crucial for segmentation tasks. This combination of residual connections, precise localization, and deep feature extraction helps improve performance in medical image segmentation tasks.

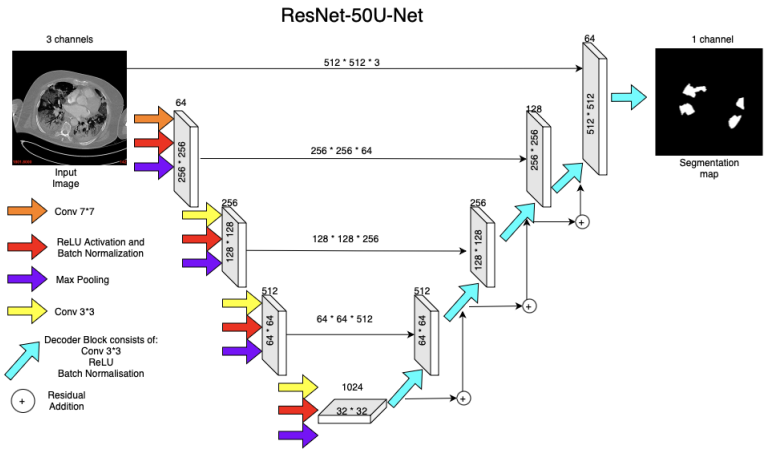


**Fig. 2.** The proposed ResNet-50U-Net Architecture

## 4 Model Training and Hyper-parameters optimisation

The tests were conducted using Google Colab environment with a T4 GPU and Visual Studio with AMD Radeon GPU. For each test, the dataset was divided into 64% of training, 16% validating and 20% test subsets. For hyper-parameters tunning, we used various loss Functions including: (a)**Dice Loss:** measures the overlap between the predicted masks and ground truth masks, (b) **Binary Cross Entropy (BCE):** measures the pixel-wise accuracy between the predicted masks and the ground truth masks, (c) **BCE Dice Loss:** combines Binary Cross-Entropy, which ensures pixel-wise accuracy, with Dice Loss, (d) **Focal Tversky Loss:** extends the Tversky index by weighting false positives and false negatives with alpha and beta, and uses gamma to focus on hard-to-segment examples, and (e)**Focal Tversky Dice Loss:** uses Dice loss for overall segmentation accuracy and the Focal Tversky loss, which emphasizes challenging regions. The *Adam optimizer* was chosen for its effectiveness in segmentation tasks. After experimentation, the ideal number of training epochs was set to 100, as higher epochs led to reduced performance on the validation set. A batch size of 2 was selected to balance memory usage and computational load. To enhance convergence, *ReduceLROnPlateau* was applied, with a learning rate of 2e-4. *Early stopping* (with a patience of 30 epochs) and *model checkpointing* were used to prevent overfitting and recover the best model weights. Lastly, *Focal Tversky Dice loss* was the best loss function, excelling in detecting smaller

GGOs. These fine-tuned parameters significantly boosted segmentation accuracy and computational efficiency.

To evaluate the optimal models generated through the random search the following metrics were used: **(a) Dice similarity score**, measures the similarity between two image samples, and it is calculated as: $\frac{2\times TP}{2\times TP+FP+FN}$, **(b) Precision**, measures the proportion of true positive predictions among all the positive predictions, and it is calculated as: $\frac{TP}{TP+FP}$ **(c) Recall**, measures the model's ability to identify all relevant instances, and it is calculated as: $\frac{TP}{TP+FN}$ **(d) Accuracy:** measures the proportion of correctly segmented pixels in relation to the total number of pixelss, and it is calculated as: $\frac{TP+TN}{TP+TN+FP+FN}$ Where TP, TN, FP, FN, denotes True Positives, True Negatives, False Positives, and False Negatives, respectively.

## 5 Results

Our experimental results are presented in Table 1. The models were evaluated on three key datasets: Train, Validation and Test and the results are summarized across the evaluation metrics mentioned above.

| Dataset | Models | Dice similarity score | Precision | Recall | Accuracy |
|---|---|---|---|---|---|
| Train | U-Net | 0.66 | 0.65 | 0.77 | 0.98 |
| | ResNet-50U-Net | 0.78 | 0.80 | 0.76 | 0.98 |
| | DenseNet-121U-Net | **0.85** | **0.83** | **0.87** | **0.99** |
| Validation | U-Net | 0.66 | 0.70 | 0.70 | 0.98 |
| | ResNet-50U-Net | 0.76 | 0.73 | **0.83** | **0.98** |
| | DenseNet-121U-Net | **0.78** | **0.75** | 0.81 | 0.98 |
| Test | U-Net | 0.67 | 0.58 | 0.81 | 0.98 |
| | ResNet-50U-Net | **0.71** | **0.63** | **0.83** | **0.98** |
| | DenseNet-121U-Net | 0.69 | 0.61 | 0.80 | 0.98 |

**Table 1.** Segmentation results obtained from the optimised U-Net, ResNet-50U-Net and DenseNet-121U-Net

Table 1 highlights that ResNet-50U-Net outperforms both standard U-Net and DenseNet-121U-Net in our GGO segmentation task. However, it is observed that the performance of DenseNet-121U-Net and ResNet-50U-Net are very close in the validation set. However, ResNet-50U-Net performs better in the test set. This is due to the ResNet50's depth and residual connections, which help capture more abstract features and improve gradient flow that is crucial for precise segmentation of intricate details. While DenseNet-121 excels in parameter efficiency (due to being less complex model), it may not perform as well in segmentation tasks demanding high-level feature separation. Furthermore, ResNet-50U-Net yeilds the highest recall value, ensuring that critical GGOs are less likely to be missed. Recall is often more important in image segmentation tasks because the primary goal is to correctly identify all the relevant regions of interest.

Figure 3 illustrates the qualitative results obtained by ResNet-50U-Net (best performing), which demonstrated superior performance in handling high-noise scans and detecting smaller GGOs.

### 5.1 Post-hoc Visual Explainability

Post-hoc visual explainability involves offering visual insights and interpretations of decisions made by DL models after their predictions are generated [12,6,2]. This process aids users in comprehending the reasoning behind a model's output which is especially vital in medical imaging, where trust and interpretability are essential. After achieving the segmentation results, GradCAM++ [4] method was applied to the best performing model, ResNet-50U-Net, to highlight important part of the image using a heatmap.

GradCAM++ is an enhanced version of GradCAM [16], which is used with CNN-based models to visualize the regions of an input image that contribute most to the model's decision. This is achieved by calculating the gradient of the target class score with respect to the final convolutional layer of the neural network. The gradient information is then used to generate a heatmap, highlighting the regions that had the most influence on the model's output. While standard GradCAM provides valuable insights, it can sometimes produce coarse and less accurate visualizations. GradCAM++ addresses these limitations by utilizing second-order gradients, resulting in a more precise and fine-grained explanation of the model's decisions. By accounting for the importance of individual pixels within the feature maps and scaling down gradients with minimal impact, GradCAM++ enhances the clarity and reliability of the visual explanations. Finally, a heatmap is generated, normalized, and amplified to visually represent the areas most influential in the model's decision-making (see Figure **??**)

While high recall value provides the clinician with confidence that a GGO has been detected, the XAI method (GradCAM++) visualises its extent via the heat map. The latter helps the clinician to explain the radiological findings to the patient, during a clinic consultation to aid treatment options in context of an informed discussion. As mentioned, GGOs have a wide range of aetiology, so further appropriate investigations can aid stratification. However, the robustness of the detection of GGOs has wider implications, allowing for the investigation of other lung disease types and their progression.

## 6 Conclusion

Ground-glass opacities are non-specific often associated with several lung disorders ranging from infection, chronic interstitial disease and acute alveolar disease to pre-invavsive lesions in CT imaging. GGO segmentation is a challenging as there is a range of morphological characteristics with low-intensity pixel contrast with adjacent structures within CT images. In this study, we analysed a set of 62 pseudonymised chest CT scans, followed by a series of data preprocessing steps and optimization procedures to identify the optimal training

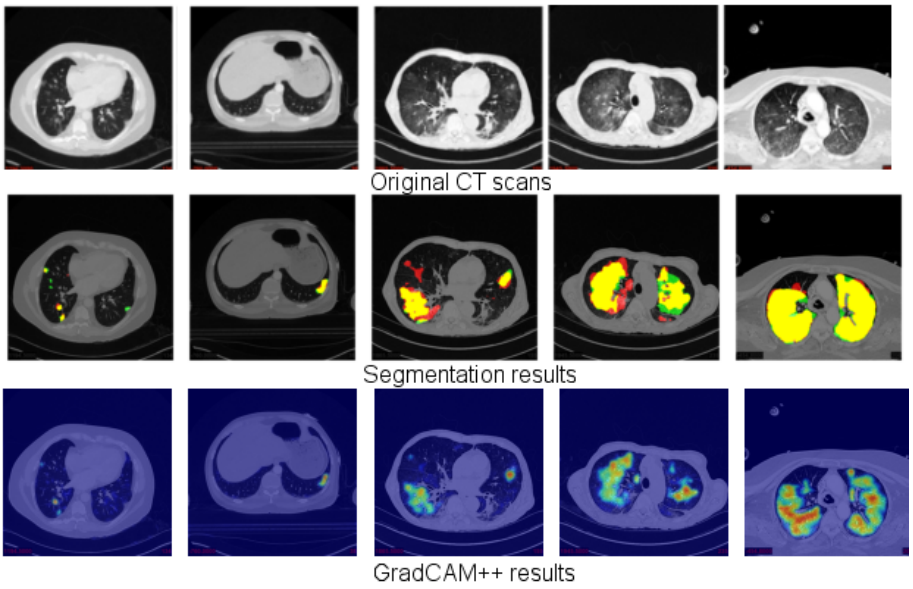


**Fig. 3.** (top row) original CT scans.(middle row) Segmentation result on six test set examples. Green represents the manual segmentation (Ground truth masks), red represents the automatic segmentation and yellow is the overlap between the automatic and manual segmentation. (bottom row) GradCAM++ results on ResNet-50U-Net model.

parameters for GGO segmentation. The segmentation results revealed that the ResNet-50U-Net model outperformed the standard U-Net and DenseNet121-U-Net. Furthermore, an explain- able framework, GradCAM++, was applied into our segmentation workflow to support clinicians in making more informed decisions by providing clearer visual evidence of the model's focus areas.

**Acknowledgement** The authors wish to acknowledge Dr Ramya DHANDAPANI, Consultant Chest Radiologist, formerly Sandwell and West Birmingham NHS Trust.

## References

1. Banday, S.A., Nahvi, R., Mir, A.H., Khan, S., AlGhamdi, A.S., Alshamrani, S.S.: Ground glass opacity detection and segmentation using ct images: an image statistics framework. IET Image Processing **16**(9), 2432–2445 (2022)
2. Barzas, K., Fouad, S., Jasa, G., Landini, G.: An explainable deep learning framework for mandibular canal segmentation from cone beam computed tomography volumes. In: The 12th International Conference on Computational Advances in Bio and Medical Sciences. Springer (2023)
3. Cao, L., Wang, K., Xing, Q., Lin, B., Zhang, Y.: Auto detection of lung ground-glass opacity nodules based on high-pass filter and gaussian mixture model. Journal of Medical Imaging and Health Informatics **6**(2), 320–327 (2016)

4. Chattopadhay, A., Sarkar, A., Howlader, P., Balasubramanian, V.N.: Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In: 2018 IEEE winter conference on applications of computer vision (WACV). pp. 839–847. IEEE (2018)
5. Fan, D.P., Zhou, T., Ji, G.P., Zhou, Y., Chen, G., Fu, H., Shen, J., Shao, L.: Inf-net: Automatic covid-19 lung infection segmentation from ct images. IEEE transactions on medical imaging **39**(8), 2626–2637 (2020)
6. Fouad, S., Hakobyan, L., Kavakli, M., Atkins, S., Bhatia, B., Rajasekaran, A., Nagori, P., Morlese, J., Fratini, A., E Ihongbe, I.: Cherie: User-centred development of an xai system for chest radiology through co-design (2024)
7. Hansell, D.M., Bankier, A.A., MacMahon, H., McLoud, T.C., Muller, N.L., Remy, J.: Fleischner society: glossary of terms for thoracic imaging. Radiology **246**(3), 697–722 (2008)
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
9. Hewitt, M.G., Miller Jr, W.T., Reilly, T.J., Simpson, S.: The relative frequencies of causes of widespread ground-glass opacity: a retrospective cohort. European Journal of Radiology **83**(10), 1970–1976 (2014)
10. Hirayama, K., Miyake, N., Lu, H., Tan, J.K., Kim, H., Tachibana, R., Hirano, Y., Kido, S.: Extraction of ggo regions from chest ct images using deep learning. In: 2017 17th International conference on control, automation and systems (ICCAS). pp. 351–355. IEEE (2017)
11. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4700–4708 (2017)
12. Ihongbe, I.E., Fouad, S., Mahmoud, T.F., Rajasekaran, A., Bhatia, B.: Evaluating explainable artificial intelligence (xai) techniques in chest radiology imaging through a human-centered lens. PLoS ONE (2024)
13. Kuo, C.F.J., Huang, C.C., Siao, J.J., Hsieh, C.W., Huy, V.Q., Ko, K.H., Hsu, H.H.: Automatic lung nodule detection system using image processing techniques in computed tomography. Biomedical Signal Processing and Control **56**, 101659 (2020)
14. Panda, S.K., Chandrasekhar, A., Gantayat, P.K., Panda, M.R.: Detecting brain tumor using image segmentation: A novel approach. In: Data Engineering and Intelligent Computing: Proceedings of 5th ICICC 2021, Volume 1, pp. 351–362. Springer (2022)
15. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. pp. 234–241. Springer (2015)
16. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision. pp. 618–626 (2017)
17. Theckedath, D., Sedamkar, R.: Detecting affect states using vgg16, resnet50 and se-resnet50 networks. SN Computer Science **1**(2), 79 (2020)
18. Ye, W., Gu, W., Guo, X., Yi, P., Meng, Y., Han, F., Yu, L., Chen, Y., Zhang, G., Wang, X.: Detection of pulmonary ground-glass opacity based on deep learning computer artificial intelligence. Biomedical engineering online **18**, 1–12 (2019)