3D Grad-CAM in Lung Cancer Images using Deep Learning Techniques

P. Bhavani[1], PL. Chithra[*]

{bhavanipandiyan.91@gmail.com[1], *Corresponding Author E-mail: chitrasp2001@yahoo.com}

Department of Computer Science, University of Madras, Chennai - 600 025, India[*]

Abstract. Medical image processing approach play a vital role in 3D Convolutional Neural Networks (CNN) using Grad-CAM (Gradient-Weighted Class Activation Mapping) techniques. This proposed 3D Grad-CAM architecture identifies the part of the tumor in the input image and utilizes the Gradient Class Activation map to classify the feature map and highlight the tumor to predict the tumor regions in the convolution last layer. This 3D network is a very difficult task to segment the regions in two-dimensional (2D) images into three dimensional (3D) images Digital Imaging and Communications in Medicine (2D DICOM). The 2D lung cancer DICOM dataset images are to A Large-Scale CT and PET/CT Dataset for Lung Cancer Diagnosis (Lung-PET-CT-Dx) taken from The Cancer Imaging Archive (TCIA) Portal. The proposed 3D DICOM Shape Conversion techniques are used to convert the 2D image dimension (x-axis of width and y-axis o height) to 3D image dimension (x-axis of width, y-axis of of height, and z-axis of depth) with the facilitate the slice and size of the lung image by applying normalization steps refine and enhance the lung images saved in 3D volume. After creating a 3D volume to perform training and testing with the proposed architecture of 3D GradCAM – CNN generates the Grad-CAM Map by combining heat map and overlay and predicting the highlight tumor region. Our experimental results achieved accuracy of 0.85, Precision of 0.90, recall of 0.82, and Fl – Score of 0.86 outperformed the results of highlighted tumor regions than pre-trained EfficientNet and ResNet (A Residual Neural Network) architectures.

Keywords: Lung Computed Tomography (CT), 3D Grad – CAM, 3D Convolutional Neural Networks, 3D Heatmap and Overlay, Deep learning.

1 Introduction

Lung cancer is an acute illness classified into two categories such as non-small cell and small-cell lung cancer, and the most prevalent cancer that results in death worldwide. It is essential to note that non-smokers are considered passive smokers and have a high chance of getting affected by lung cancer. It is crucial to detect, prevent, and improve patient treatment early on. Cancer Imaging techniques, namely MRI (Magnetic Resonance Imaging), PET scan (Positron Emission Tomography), and CT scan (Computed Tomography) can aid in the staging and diagnosis of lung cancer for patients so they will receive chemotherapy and radiation in the future [1].

Imaging methods, including computed tomography of chest and chest X-ray, have been employed extensively pulmonary lesions to be evaluated. Even X-ray scanners indicate improved

accessibility, most lung cancer sufferers exhibit ground-class of bilateral pulmonary parenchymal opacities and lung combination with a around morphology [3], making the distinction between lung cancer infection and chest X-rays difficult. On the other hand, 3D chest CT is successfully able to distinguish between soft tissue and visualize the morphological patterns of pulmonary parenchyma. It has been widely used for the diagnosis of lung cancer and is considered as a valuable complement to rRT-PCR (reverse transcription polymerase chain reaction) experiments. [2], [4]. Radiologists are greatly helped by computer-assisted interpretation and automatic lung region-of-interest (ROI) extraction to diagnose and treat patients [5] accurately.

Deep learning's accurate and reliable performance makes it an essential tool for medical image processing [6]. The lung cancer diagnosis of chest CT has been the focus of some researchers' studies in computer vision using deep learning. Contrary to most currently utilized Artificial Intelligence algorithms in lung cancer diagnosis, which is centered on the goal of end-to-end classification process is to identify the clinic interpretable knowledge by extracting and analyzing the image features.

The proposed analysis uses convolutional neural networks (CNN) and deep learning to provide the radiologist with automatic 3D diagnosis and treatment techniques. The feature heatmaps and Grad-CAM (Gradient-Weighted Class Activation Mapping) were visualized to demonstrate in the lung region with system's dependability. Furthermore, the proposed system differentiates more clearly between the highly similar lung cancer types adenocarcinomas, squamous carcinomas cell, and large cell carcinomas due to this technique, radiologists can quickly identify the tumors, and it will help doctors to treat the patients efficiently.

This research work consists of four sections: Section 2 describes related works on 3D Grad-CAM lung cancer highlighting regions; Section 3 describes the proposed 3D Grad-CAM Convolutional Neural Network Layer architecture; discusses the experimental results and their comparisons in Section 4; and Section 5 generates conclusions.

2 Related Work

While there is a lot of related work on automated medical image segmentation, this novelty focused only on CNN-based algorithms and Grad-CAM techniques, which are directly pertinent to this proposed work.

Zeiler and Fergus created the deconvolution strategy to comprehend the upper levels in a specific network to make a dataflow of neuron activation from top most layers to the bottom of the layer in the image. This method highlights the tumor regions of the lung image that most powerfully stimulate that improved neuron [1].

This work was expanded by Springenberg et al. [2] to a novel methodology termed guided backpropagation, which, in contrast to existing visualization techniques, enabled comprehension of how each deep neural network neuron affects the input lung image [3].

Yosinski et al. proposed a method for generating an image as input that induces a particular region within a neural network to be highly activated to visualize the unit's functionality [4].

Wu, Xiangjun et al. provided that on the ImageNet datasets, several of the most potent EfficientNet architectures had the highest precision. They are included in this experiment to investigate contrast and the possibility of performance enhancement [5], [6].

Yousefzadeh et al. developed medical image tools and artificial intelligence in a deep-learning framework for EfficientNetB3-based feature extraction. This framework diagnoses lung tumors using CAM in COVID chest CT scans MosMedData cohort dataset [7].

Hara, K., Kojima et al. implement the class activation mapping of gradient-based to predict the regions correctly and defects can analyze the crystal growth development of a better fabrication process [8]. The Grad-CAM(Gradient-weighted Class Activation Mapping) is used to classify the regions using CNN models in different architectures VGG and ResNet. The convolutional layer creates a high-resolution tentative localization map highlighting and predicting the most critical regions in the lung image using discriminative of class visualization, Guided Grad-CAM [9].

Bau et al. Proposed textual justifications for model decisions in Computer vision and pattern recognition. Ultimately, we plan and carry out human experiments to ascertain if Grad-CAM explanations help users put the right amount of trust in deep network analyses. Based on our findings, novice users can effectively distinguish between a "stronger" and a "weaker" deep network using Grad-CAM, even when they produce identical predictions [10].

Shangdong Zhu et al. propose the GW-Net, Re-ID network is one of powerful semi-supervised with Grad-CAM consistency standardize and deterioration of randomly erasing characteristics. The primary goal is to record in-depth data and anticipate effective performance under cross-domain conditions [12].

The following are the primary benefits and novelty of the present research:

- Converting 2D volume DICOM image to 3D DICOM image dimension to create the dataset to train and test the proposed 3D GradCAM architecture.

- Proposing 3D GradCAM architecture, which is used to detect and diagnose, classifies tumor regions accurately.

- The final convolutional layer of Grad-CAM is highlighting and predicting the tumor region.

- Combining and evaluating the Grad-CAM, Heatmap, and overlay diagnosis is helpful in predicting tumor regions accurately.

This proposed 3D GradCAM architecture results where compared with pre-trained EfficientNet and ResNet (A Residual Neural Network) architectures.

3 Proposed Methodology

The proposed architecture of the 3D GradCAM network is used to classify tumor and non-tumor regions. To extract the tumor regions and process of identifying lung tissues performing with GradCAM and overlay techniques. The proposed workflow is conferred in Figure 1.
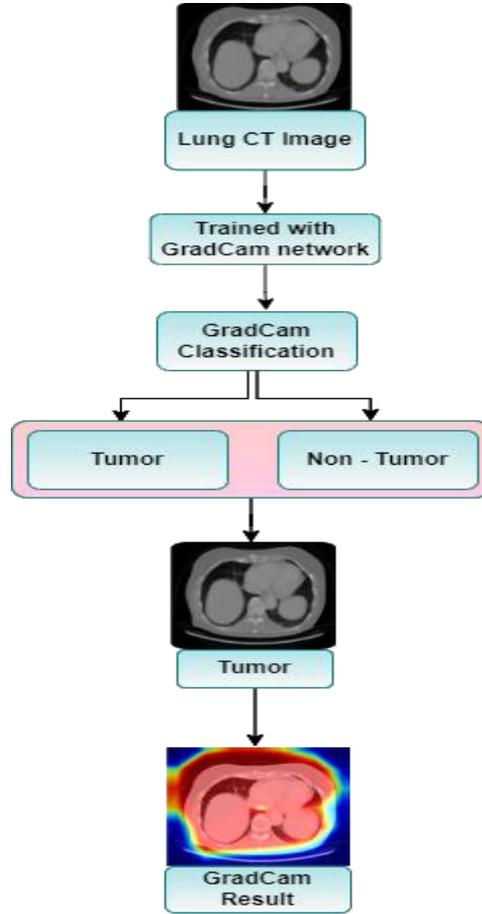
Fig. 1. 3D Grad-CAM Proposed Methodology Workflow

3.1 LungCancer Dataset

The Proposed 3D Grad -CAM architecture dataset can be taken from TCIA (Cancer Imaging Archive) data repository in Lung-PET-CT-Dx (Large-Scale CT and PET/CT Dataset for Lung Cancer Diagnosis) that includes 175 CT images in the format of DICOM with 64 slices in each folder and a tumor size of 5mm [11]. The dataset can be classified into 2 categories, which include 6300 images of 175 people with lung cancer (36 images/person) and 4900 lung images of 175 normal patients (~28 lung images/person), as shown in Table 1. All grayscale lung images are collected from patients and made freely accessible through the TCIA data portal.

Table 1. Lungcancer Dataset

Tumor stage	Lung CT Dataset		
	Patients	Lung Images	*Slices of each patient*
Normal	175	4900	64
Abnormal	175	6300	64

3.2 Proposed 3D DICOM Shape Conversion

The Medical image is a 2D DICOM image format containing information, location, and tumor position in each slice, as shown in Figure 2. This slice consists of an x and y-axis. This proposed method arranges the DICOM slices in the correct manner and predicts the image size and number of slices in the DICOM folder. Each input Slice indicates the tumor size and position of the tumor between slices along with the z-axis to visualize and generate a 3D Dicom Image is defined in equation (1). Finally, the 2D DICOM is reconstructed into the 3D dataset and saved in 3D DICOM format. Where, 2D DICOM Dimension size (2Di), No. of Slices in 2D DICOM (NDi) as plotted in Figure 2.
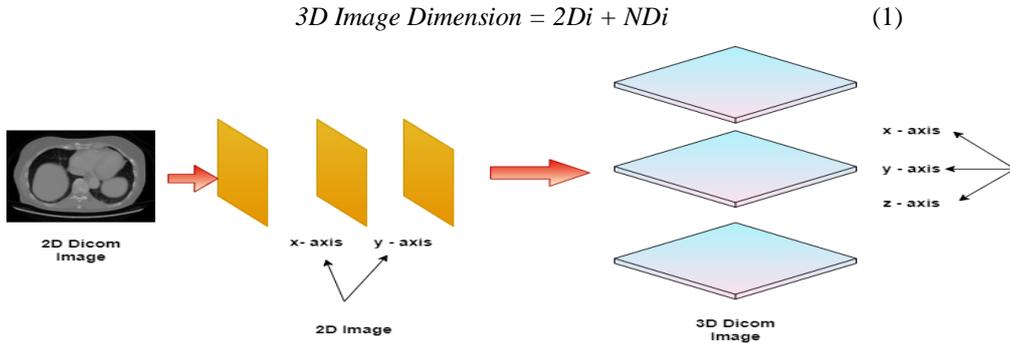
$$3D\ Image\ Dimension = 2Di + NDi \qquad (1)$$

Fig. 2. Conversion of 2D DICOM Image Dimension to 3D DICOM Image Dimension

3.3 3D Lung Cancer Pre-Processing

Data pre-processing is a crucial phase in image processing algorithms since a model develops its pattern recognition abilities depending on the receives data. An image-wise split would be preferable to a patient-wise split in order to prevent the leakage of data in the dataset training to the testing dataset. Cross-validation was used to divide all images into training and testing sets. Then, the training set and test set for each images. This prevented the same patient's images from both appearing in the training and testing sets, eliminating any model confusion arising from a learning of CNN, shape or structure of lung patients. The proposed pre-processing steps are used to reduce the image size, use the Sobel operator to enhance the tumor and normalize the pixel values as mentioned in equation (2).

$$Pixel\ Value\ (PX) = (PXi – PXi\_Min) / (PXi\_Max – PXi\_Min) \qquad (2)$$

Where the original image pixel value (PXi), Pixel Value of Minimum (PXi_Min), and Pixel Value of Maximum Value (PXi_Max).

3.4 Proposed 3D Grad – CAM Convolutional Neural Network

The proposed 3D Grad-CAM CNN architecture is implemented in Figure 3. This proposed network consists of eight convolutional layers and eight MaxPooling3D layers with new input each layer of shape and size of the filters are altered. The lung input shape of the 3D image (512, 512, 64) to the Proposed Grad -CAM convolutional neural networks, images size of 32 ($3 \times 3 \times 3$) (width, height, depth, number of channels). The layer of first 3D input has 32 sizes of filter with a size of kernel is $3 \times 3 \times 3$. The 3D input shape determined the input size of filter and the number of filters between the first 3D layer of convolution and the $2 \times 2 \times 2$ size of MaxPooling 3D layer with 2 strides. Reducing the data size in MaxPooling. The layer of second convolution

layer includes 64 filters and same size $3 \times 3 \times 3$ filters in the first convolutional layers in between the convolution second 3D layer, the same size of $2 \times 2 \times 2$ the preceding Maxpooling 3D layer. The following 3D convolution third layer and fourth layer of convolutional 3D layers have size $3 \times 3 \times 3$ with 128 sizes of filter. After the 3D layer of fourth convolution layer and 3D MaxPooling of the $2 \times 2 \times 2$ same size, similar to the preceding 3D MaxPooling layers. The last two-layer 3D convolution layers are used, where 256 size of filter $1 \times 1 \times 1$, and the $1 \times 1 \times 1$ with 512 sizes of filter. After, there is a final MaxPooling3D size of layer $2 \times 2 \times 2$ with 256 of 3D layers and filter of 512. Each 3D MaxPooling layer has a normalization of batch running to the background. There is a thick, flattened layer at the network's termination. There is only one dense layer, which takes the number 11 as-is. The filter sizes were selected to consider the output of filters required to follow the output. The proposed architecture performs the optimization method "Adam" and used a default learning rate of 0.1. The output for various batch size and epochs saved weight and performed the GradCAM model in the convolution last layer.
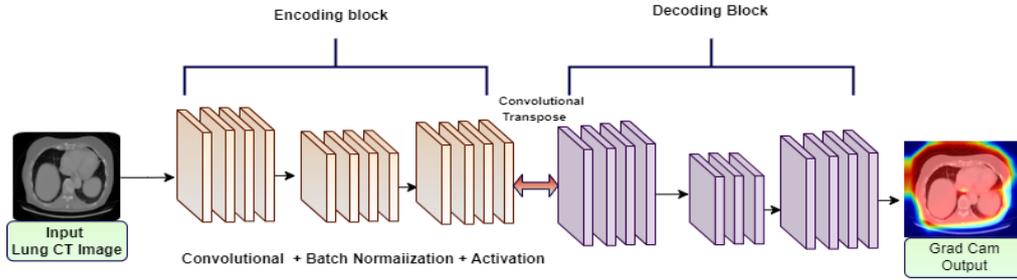

Fig. 3. Proposed 3D Convolutional Neural Network Model

3.5 Proposed 3D Grad – CAM Visualization

GradCAM visualization techniques were used to compare Lung Cancer positive and Lung Cancer negative patients. The proposed network model highlights region-based features recognized for making high-accuracy classifications. These images are generated for the best-performing models and compared with Efficient Net and ResNet50. When visualizing non-lung cancer images in a 3D model, the resulting heat maps tended to become more diffused and dispersed, even if they were still particular and localized when visualizing lung cancer.

The CAM [7][9]and Grad-CAM[8] Methods are founded on the essential premise that the score of final Y^c for a particular c of class can be expressed as an ordered sequence of its average global pooling last layer of convolutional feature mappings A^k.

$$Y^c = \sum_k w_k^c \sum_i \sum_j A_{ij}^k \qquad (3)$$

In the particular class restoration map Lc, each (i, j) class location is thus determined as follows:

$$L_{ij}^c = \sum_k w_k^c . A_{ij}^k \qquad (4)$$

Each model performed cross-validation training, testing, and validation. Then, each class's accuracy, precision, recall, and F1 scores were determined. After computing the metrics' average value across all rounds, their anticipated values were provided with a 95% confidence interval. The metrics are described below.

$$\text{Accuracy} = \frac{(ATP+ATN)}{(ATP+ATN+AFP+AFN)} \qquad (5)$$

$$Precision = \frac{ATP}{(ATP+AFP)} \qquad (6)$$

$$Recall = \frac{ATP}{(ATP+AFN)} \qquad (7)$$

$$\text{F1 score} = \frac{2*(Precision\ X\ Recall)}{(Precision + Recall)} \qquad (8)$$

Where ATP (True Positive) refers to the precise count of positively detected pixels. The number of negative pixels discovered is known as ATN (True Negative). A group of positively determined positive pixels is referred to as AFP (False Positives). Incorrectly calculated negative pixels are counted as AFN (False Negative).

3.6 HeatMap Overlay

The proposed Grad-CAM, a heatmap is a type of visualization that shows the intensity with which data is present at various locations. A colored overlay will be visible on top of the map when the Heatmap Layer is enabled. By default, locations with higher intensities will be red, while those with lower intensities will be green.

4 Experimental Results and Discussion

4.1 Performance of the Proposed 3D DICOM Shape Conversion

This research work takes 11,200 lung cancer positive and negative images and the positive lung cancer tumor size is 5mm of the 2D DICOM image shape (512, 512) dataset has been converted into the 3D data (512, 512, 64) with respect to input shapes and number of slices as displayed in Figure 4.
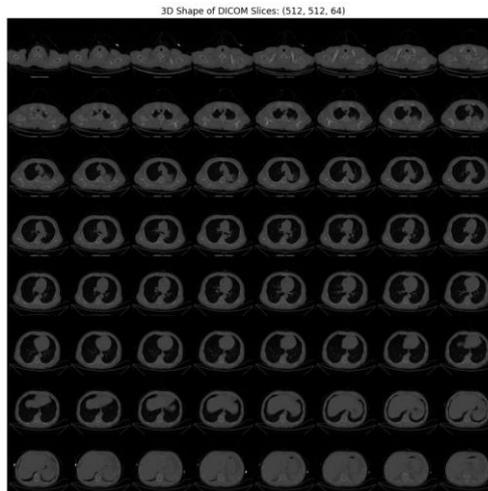

Fig. 4. Output of 3D DICOM Shape Conversion

4.2 Performance of the 3D Pre-Processing

The 3D pre-processing is a Gateway to enhance the image quality using Sobel edge detection as generated in Figure 5(a), image normalization, and rotating the tumor in different angles using data augmentation as implemented in Figure 5 (b)
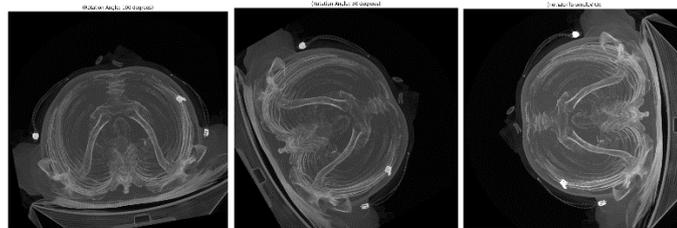

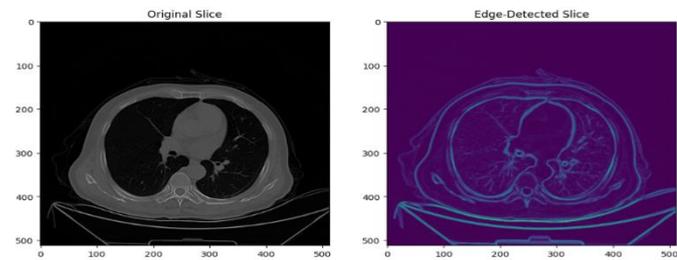**Fig. 5(a).** Output of 3D Pre-Processing (Data Augmentation)

Fig. 5(b). Output of 3D Pre-Processing (Sobel Edge Detection)

4.3 Performance of the 3D Grad CAM HeatMap and overlay

The performance of the 3D Grad-CAM CNN model compared with an efficient net and Resnet50 model. From this comparative analysis, the proposed model has achieved better results, as shown in Figure 6 (a), the output of the Efficient Net model, Figure 6(b), the output of the ResNet model, and Figure 7, as shown in the 3D Grad-CAM results.
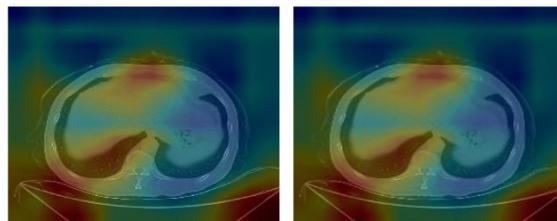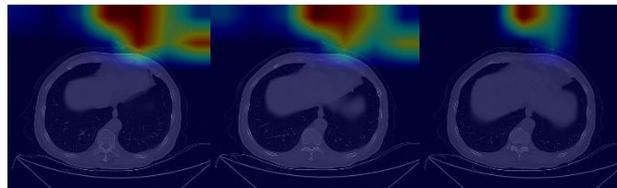

Fig. 6(a). Output of EfficientNet

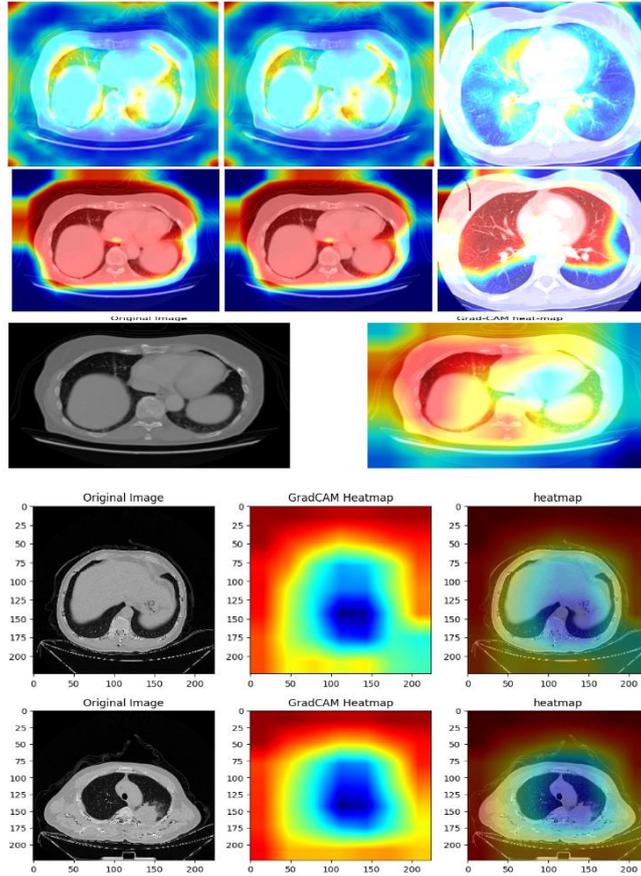**Fig. 6 (b).** Output of ResNet50

Fig. 7. Output of Grad-CAM CNN

5 Conclusion

This research focused on highlighting the lung tumor regions using the proposed 3D Grad-CAM CNN to detect and diagnose lung tumor regions. The gradient-based 3D Convolutional neural networks were used to extract the features from visual lung CT cancer images using class activation mapping and heatmap techniques were used to apply in the 3D Grad-CAM model. This method performs better in terms of efficiency and removing human bias while analyzing the tumor structure than the Efficient Net and ResNet50 models. However, tumor locations outside of the given lung image's midpoint were frequently more interesting to distinguish than the midpoint pixels. It can be observed through an in-depth analysis and prediction of highlighted lung images. This shows that an automated learning model successfully identifies lung patterns in the CT image. Future research goal at developing in depth knowledge and understanding of how lung tumour regions are generated as 3D models using huge datasets from different medical image datsets.

Acknowledgments

References

[1] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in European conference on computer vision. Springer, 2014, pp. 818–833.

[2] W. Wang, Y. Xu, R. Gao, R. Lu, K. Han, G. Wu, W. Tan, "Detection of SARS-CoV-2 in different types of clinical specimens", JAMA 323 (18) (2020) 1843–1844

[3] M. Chung, A. Bernheim, X. Mei, N. Zhang, M. Huang, X. Zeng, J. Cui, W. Xu, Y. Yang, Z.A. Fayad, "CT imaging features of 2019 novel coronavirus (2019-nCoV)", Radiology 295 (1) (2020) 202–207.

[4] L.J.M. Kroft, L. van der Velden, I.H. Girón, J.J.H. Roelofs, A. de Roos, J. Geleijns, "Added value of ultra–low-dose computed tomography, dose Equivalent to chest x-ray radiography, for diagnosing chest pathology", J. Thorac. Imaging 34 (3) (2019) 179–186.

[5] Wu, Xiangjun, Hui, Hui, Niu, Meng, Li, Liang, Wang, Li, He, Bingxi, Yang, Xin, Li, Li, Li, Hongjun, Tian, Jie, & Zha, Yunfei (2020). "Deep learning-based multi-view fusion model for screening 2019 novel coronavirus pneumonia: A multicentre study". European Journal of Radiology, 128, 109041. https://doi.org/10.1016/j. ejrad.2020.109041.

[6] Z. Hu, J. Tang, Z. Wang, K. Zhang, L. Zhang, Q. Sun, "Deep learning for image-based cancer detection and diagnosis-a survey", Pattern Recognit. 83 (2018) 134–149.

[7] Yousefzadeh, M., Esfahanian, P., Movahed, S. M. S., Gorgin, S., Lashgari, R., Rahmati, D., Kiani, A., Kahkouee, S., Nadji, S. A., Haseli, S., Hoseinyazdi, M., Roshandel, J., Bandegani, N., Danesh, A., Bakhshayesh Karam, M., & Abedini, A. (2020). "ai-corona: Radiologist-Assistant Deep Learning Framework for COVID-19 Diagnosis in Chest CT Scans". https://doi.org/10.1101/2020.05.04.20082081.

[8] Hara, K., Kojima, T., Kutsukake, K., Kudo, H., & Usami, N. (2023). "3D CNN and grad-CAM based visualization for predicting generation of dislocation clusters in multicrystalline silicon. APL Machine Learning", 1(3), https://doi.org/10.1063/5.0156044.

[9] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, "Grad-cam: Why did you say that? visual explanations from deep networks via gradient-based localization," arXiv preprint arXiv:1610.02391, 2016.

[10] Roy, P. K., & Kumar, A. (2022). "Early prediction of COVID-19 using ensemble of transfer learning". Comput. Electr. Eng., Article 108018. https://doi.org/10. 1016/j.compeleceng.2022.108018. https://www.sciencedirect.com/science/article/ pii/S004579062200283X.

[11] The TCIA DataSet, https://www.cancerimagingarchive.net/

[12] Shangdong Zhu, Yunzhou Zhang, Yu Feng, GW-net: "An efficient grad-CAM consistency neural network with weakening of random erasing features for semi-supervised person re-identification, Image and Vision Computing", Volume 137, 2023, 104790, ISSN 0262-8856, https://doi.org/10.1016/j.imavis.2023.104790.