International Journal of Innovative Research in Computer Science and Technology (IJIRCST)
ISSN(Online): 2347-5552, Volume-12, Issue-3, May 2024
https://doi.org/10.55524/ijircst.2024.12.3.16
Article ID JIRD-1297, Pages 100-105
www.ijircst.org

Deep Learning-Based Lung Medical Image Recognition

Xinghui Fei[1], Yulu Wang[2] , Lu Dai[3], and Mingxiu Sui[4]

[1]Colorado State University, Fort Collins, USA
[2]Ottawa University, Phoenix, USA
[3]University of California, Berkeley, Berkeley, USA
[4]University of Iowa, Iowa, USA

Correspondence should be addressed to Sheng Chai; xinghui2023@gmail.com

Received: 25 April 2024 Revised: 10 May 2024 Accepted: 23 May 2024

Copyright © 2024 Made Sheng Chai et al. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT- Pulmonary nodules serve as critical indicators for early lung cancer diagnosis, making their detection and classification essential. The prevalent use of transfer learning in recognition algorithms often encounters a significant disparity between source and target datasets, which hampers effective feature extraction from pulmonary nodules and degrades performance. An enhanced neural network model leveraging convolutional neural networks is introduced to address this issue. This model integrates a pre-trained GoogLeNet Inception V3 network with a custom-designed feature fusion layer, improving the network's ability to extract features. To ascertain the optimal configuration, the models were evaluated based on accuracy in various combinations. The experiments conducted on the LUNA16 pulmonary nodule dataset revealed that the refined network model achieved an accuracy of 88.78% and a sensitivity of 87.18%. This represents an increase of 2.7 and 2.22 percentage points in accuracy and sensitivity, respectively, compared to the GoogLeNet Inception V3 algorithm. Further tests across different dataset proportions also yielded superior outcomes, demonstrating enhanced generalization capabilities. These findings can offer objective benchmarks for clinical diagnosis.

KEYWORDS- Deep Learning, Target Detection, Neural Network, Medical Image, Transfer Learning, GoogLeNet

I. INTRODUCTION

The incidence and mortality rates of lung cancer continue to rise each year. According to the American Cancer Society, lung cancer accounts for over 14% of all new cancer diagnoses annually. Lung cancer leads in both incidence and mortality among all malignant tumors. Pulmonary nodules, which are often multi-organ and multi-system diseases linked to lung cancer, could be early markers of the disease. Consequently, detecting pulmonary nodules is critically important for the early diagnosis of lung cancer [1-2].

Computed Tomography (CT), a prevalent imaging technique for evaluating pulmonary nodules, necessitates that physicians possess specialized medical imaging knowledge and clinical experience. Additionally, the variability in diagnostic outcomes among physicians [3] has led to the widespread adoption of Computer-Aided Diagnosis (CAD) systems in recent years. These systems offer relatively objective metrics to support clinical decision-making.

Traditionally, research methods involved labor-intensive feature extraction from images, identifying regions of interest (ROI), and analyzing nodules based on shape, grayscale, and texture. However, with the swift advancement in the field of machine learning[4-5], deep learning algorithms like convolutional neural networks have become commonplace in various areas[6-10]. These include two-stage detection algorithms focused on candidate regions such as RCNN, and regression-based one-stage detection algorithms like the YOLO [11] series. Researchers have applied the YOLO framework to detect nodules in CT scans using CNNs to predict multiple bounding boxes, achieving a sensitivity of 89% on the LIDC-IDRI dataset[12]. Commonly, networks such as CNN, YOLO, FCN, and R-CNN trained on different datasets are used to extract features from the target dataset [13-16].

Despite the successes of transfer learning, the significant disparity between source and target datasets suggests potential improvements in feature extraction. To address this, we propose an enhanced GoogLeNet Inception V3 deep neural network model for pulmonary nodule detection. The core of this work involves developing a feature fusion layer and optimizing relevant parameters to enhance feature extraction from the target dataset.

II. RELATED WORK AND METHODOLOGY

A. Overview of GoogLeNet

Szegedy and colleagues [17] introduced the GoogLeNet network, which represented a significant advancement in the field of deep learning. Distinct from earlier architectures like AlexNet and VGG, GoogLeNet featured a more profound and complex architecture yet utilized fewer parameters, as highlighted in Table 1. This network redesign, detailed in Figure 1, incorporated the innovative Inception layer. The Inception layer, a novel concept at the time, allowed for a substantial reduction in the number of parameters without sacrificing the network's performance. This design enabled more efficient computation and training processes, making GoogLeNet not only a powerful tool for image recognition tasks but also a more practical model for varying hardware capabilities. The introduction of auxiliary classifiers to enhance gradient flow at deeper layers and the use of a global average pooling layer instead of fully connected layers at the top also contributed to its efficiency and effectiveness. These

features helped in mitigating the vanishing gradient problem, which is common in deeper networks, thus ensuring better convergence during training.

Table 1: Comparison of commonly used neural network parameters

Model	TOP-1 Accuracy/%	Number of Parameters	Depth
Inception V3	78.8	23 851 773	153
Resnet50	75.7	25 636 701	164
VGG16	71.3	138 357 533	25
VGG19	72.5	143 667 232	26

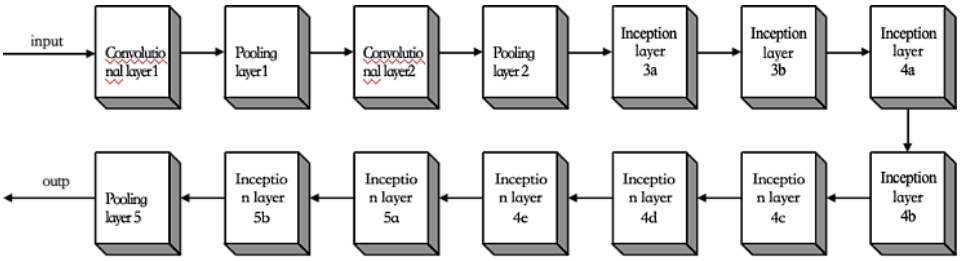

Figure 1: Network structure diagram of GoogLeNet

B. Enhanced RGIV3 Network Design

The GoogLeNet Inception V3 network [18], originally trained on the ImageNet dataset, shows a considerable disparity when applied to the LUNA16 dataset utilized in this study, leading to suboptimal performance. To address this, a fusion layer was designed to mitigate the effect of this discrepancy on the recognition process. This fusion layer facilitates the nonlinear fitting of network-learned feature information[19-20], enhancing the extraction of image features.

The fusion layer is composed of three fully connected layers and one dropout layer, as depicted in Figure 2. This layer integrates with the GoogLeNet Inception V3 network, hereinafter referred to as GIV3, with the integrated network named RGIV3.

Based on empirical data, node numbers should be selected from 1024, 512, 256, 128, with dropout rates between 0.4, 0.5, 0.6. To identify the optimal setup, the study tests various combinations of node numbers and dropout rates in 30 experimental groups, using accuracy as the criterion for selection. As shown in Table 2, the chosen parameter combination is 1024 and 512, with a dropout rate of 0.5.

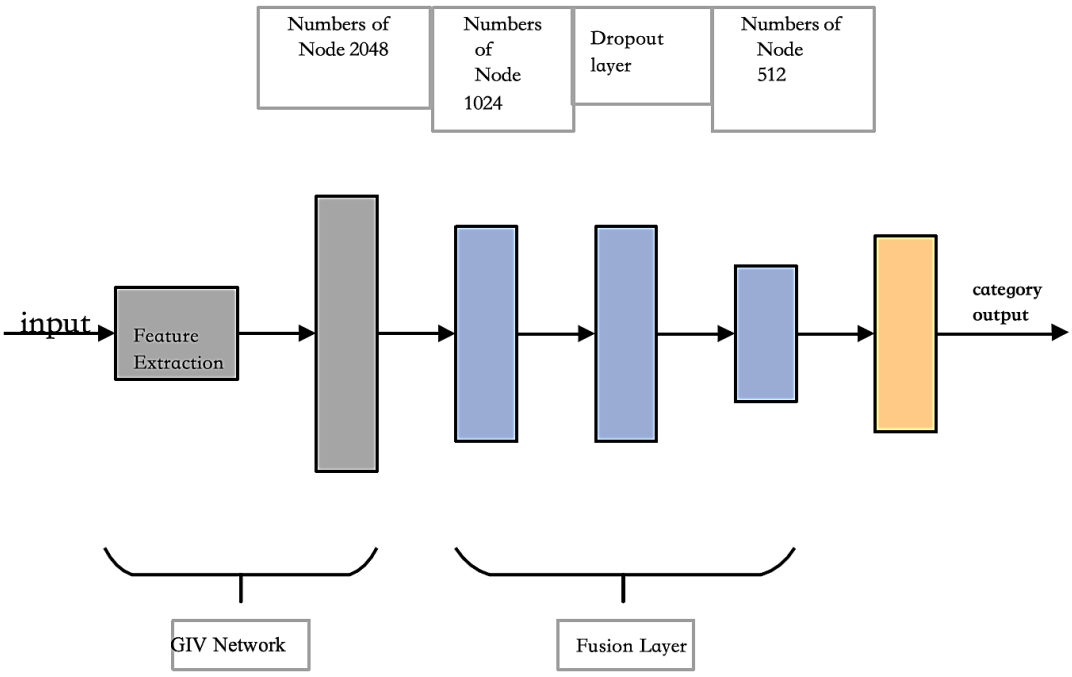

Figure 2: Improved RGIV3 structure diagram

Table 2: Experimental Results of Node Count and Dropout Rate Combinations

Node Number Combination	Dropout rate 0.3	Dropout rate 0.4	Dropout rate 0.5
1 024, 1024	86.33	87.64	88.62
1 024, 512	87.71	88.44	88.78
1 024, 256	86.53	87.61	88.56
1 024, 128	87.26	86.81	88.32
512, 512	86.26	84.41	87.32
512, 256	86.26	87.13	88.10
512, 128	86.76	87.10	86.83
256, 256	81.41	87.46	88.21
256, 128	83.03	84.51	87.26
128, 128	86.26	84.43	87.22

Given the limited sample size of the target dataset, there is a high risk of overfitting when training a new model. To address this, the experiments utilize weights pre-trained on large-scale databases like ImageNet as the initial basis. Fine-tuning is then performed at a lower learning rate to optimize performance.

C. Methods for Updating Network Weights

The selected method for updating the network weights is Stochastic Gradient Descent (SGD)[21-22], which involves updating the weights in the direction that reduces the learning rate towards the minimum of the loss function. The formula is detailed below:

$$W_{i+1} = W_i - \Delta\text{his}_{i+1} \quad (1)$$

In this formula, $W_{i\square 1}$ denotes the weights after the i+1 iteration; W_i denotes the weights after the *i* iteration; $\square\text{his}_{i\square 1}$ indicates the accumulated gradient. The specific formula is presented as

$$\Delta\text{his}_{i+1} = \text{momentum} \times \Delta\text{his}_i + L \times \frac{\partial \text{loss}}{\partial w} \quad (2)$$

Here, Δhis_{i+1} represents the accumulated gradient after the *i+1* iteration; Δhis_i represents the accumulated gradient after the *i* iteration; 'momentum' refers to the momentum coefficient; *L* represents the current learning rate; $\frac{\partial loss}{\partial w}$ denotes the derivative of the loss function.

III. EXPERIMENTAL DESIGN

A. Dataset

This study utilizes the LUNA16 dataset, which serves as a subset of the publicly accessible lung nodule dataset LIDC-IDRI[23]. It comprises CT images from 888 patients, divided into 10 subsets for storage. Each image set includes an mhd file and a raw file; the mhd file contains essential image metadata, while the raw file contains the pixel data. An illustrative example of a CT image is depicted in Figure 3.

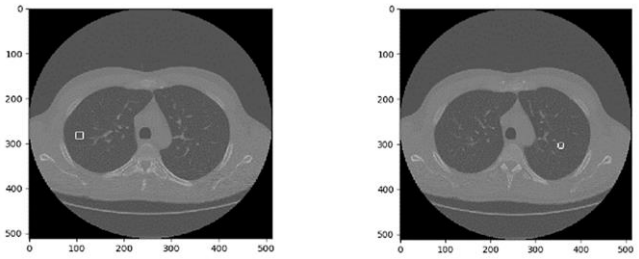

Figure 3: Example of a LUNA16 Image

B. Preprocessing of the Dataset

The LUNA16 dataset comprises raw image data (.raw) and metadata annotations (.mhd). Neural networks are unable to directly interpret these formats, The dataset was handled using the approach of Linked Data, which consolidates a variety of data formatations, a key aspect in academic studies [24]. This is well-organized handling of the data set that facilitates cross-referencing the data, hence enhancing interoperation among other data sets. This particularly comes in handy in areas such as machine learning and artificial intelligence, where data quality is everything in effective model training for accurate results. thus, images in .mhd format are converted to the universally compatible BMP format. This conversion yields RGB images equivalent to the number of CT slices, each displaying the complete information of a single slice. The annotation files provide the locations and characteristics of nodules, which, combined with slice depth, facilitate rapid localization of nodules and corresponding slices [25]. Variations in patient positioning during CT scans can cause image orientation issues, necessitating adjustments based on annotation data to reorient images correctly and update coordinate information. The small size of the dataset requires augmentation to ensure adequate training data volume. This study employs supervised data augmentation techniques on the BMP images to enrich the dataset and utilizes fine-tuning of the learning rate to mitigate overfitting.

The pulmonary parenchyma includes the bronchial and terminal alveolar structures, representing the air-contacting cavities and walls within the lungs. In CT imaging, this is indicated by the black areas within the lung's white outlines. The focus of this research is on detecting pulmonary nodules, hence only the pulmonary parenchyma is analyzed. External contours of the lungs, which can disrupt neural network training, are excluded from analysis following image segmentation to isolate the pulmonary parenchyma.

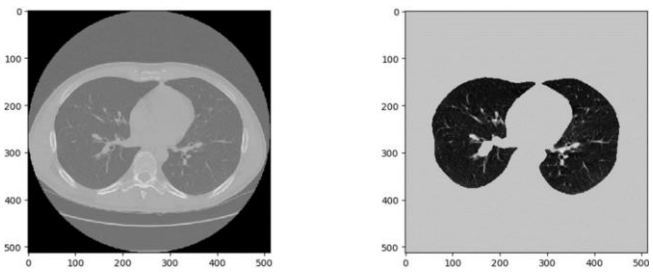

Figure 4: Original Image and Processed Pulmonary Parenchyma Image

Figure 4 displays a typical slice from the LUNA16 dataset. Following binary transformation and morphological

operations like erosion and dilation, the pulmonary parenchyma is delineated from the lung contours as illustrated in the right-hand image. Subsequent processes focus solely on the isolated pulmonary parenchyma.

C. Experimental Procedure

The experiments were executed on a system running Windows 10, equipped with an Intel Core i5-6200U processor at 2.30 GHz, and utilized the TensorFlow framework and slim for fine-tuning the model, with Python version 3.6.

The experimental setup included training on datasets from 740 patients and testing on datasets from 130 patients. Initially, 2,000 images each of lung nodules and healthy tissues were extracted. Data augmentation techniques expanded the dataset to 5,000 images each of lung nodules and healthy tissues. Images were resized to a uniform 330×330 pixels. Data splits for training and testing were varied as follows: 80% training and 20% testing, 70% training and 30% testing, and 60% training and 40% testing. Each configuration was trained until the loss function stabilized, followed by performance testing. Specifically, the 80% training split underwent 2775 training steps on the slim model, reducing the loss function to 0.37 and achieving an accuracy of 88.79%; results for other splits are detailed in Table4.

In these experiments, the bespoke RGIV3 network underwent training on both the training and validation sets and was evaluated on the test set. To circumvent the risk of converging to local optima, a step-wise reduction strategy was employed for the learning rate, adjusting it every 200 iterations.

D. Analysis of Experimental Results

The results of the experiments provided measurements for True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN). For assessing the performance of the model, three key metrics were selected [26]: Accuracy (Accu), Sensitivity (Sensi), and Specificity (Speci). The formulas used to compute these metrics are:

$$\text{Accu} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

$$\text{Sensi} = \frac{TP}{TP+FN} \quad (4)$$

$$\text{Speci} = \frac{TN}{TN+FP} \quad (5)$$

Accuracy measures the degree of concordance between predicted and actual outcomes, while sensitivity assesses the model's capability to detect pulmonary nodules. Specificity, on the other hand, evaluates the model's ability to identify images that do not contain pulmonary nodules. The training processes for GIV3 and RGIV3 are depicted in Figure 5. As observed from Figure 5, RGIV3 demonstrates quicker convergence and enhanced stability in performance. Moreover, when considering accuracy, the enhanced RGIV3 model has shown superior performance on the dataset.

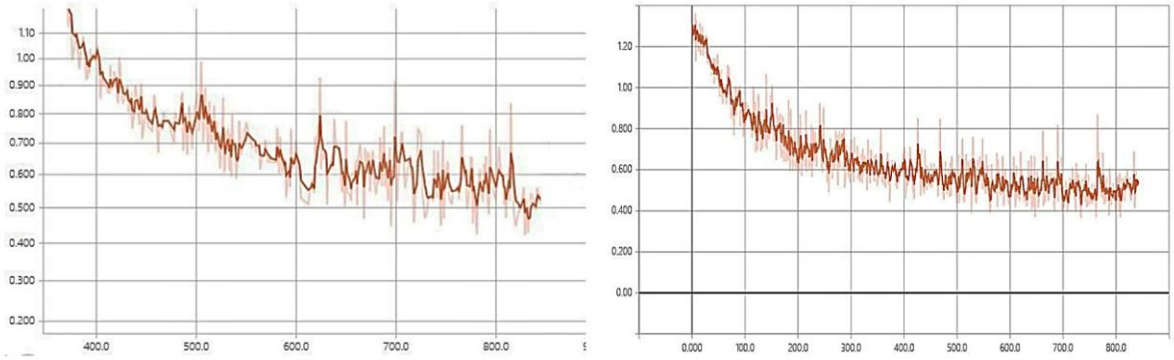

Figure 4: Training Process (Variation in Overall Loss)

When utilizing 80% of the data as the training set and 20% as the test set, the comparative performance of each model is illustrated in Table 3.

Table 3: Comparison of Model Performance

Experiment NumberAccuracy	Recognition Model	Accuracy	Sensitivity	Specificity
1	Alexnet	85.35	87.73	73.0
2	Yolo V3	84.05	88.75	82.2
3	GIV3	86.06	85.01	80.5
4	RGIV3	88.78	87.18	80.6

Table 3 indicates that the enhanced model presented in this paper has achieved accuracy and sensitivity rates of 88.78% and 87.18%, respectively. These represent increases of 2.7% and 2.22% compared to the traditional GIV model, thereby achieving superior recognition performance.

To further assess the improved RGIV3 model comprehensively, experiments were carried out using various training-to-testing set ratios, with the outcomes detailed in Table 4.

Table 4: Model performance comparison under different data ratios

The ratio of training set to test set	Accuracy Of GIV3	Accuracy Of RGIV3
8-2	86.07	88.79
7-3	84.20	87.56
6-4	84.84	87.22

Based on the above table and considering the training process analysis, the improved RGIV3 model exhibits better

recognition performance compared to the original GIV model. It outperforms GIV3 in recognition effectiveness at data ratios of 80%, 70%, and 60%. Additionally, it converges faster, indicating superior performance.

IV. CONCLUSION

This study presents a novel approach to detecting pulmonary nodules by utilizing an enhanced neural network model trained on a subset of the LUNA16 dataset. The integration of the GIV3 network with a meticulously crafted feature fusion layer augments the model's proficiency in feature extraction, thereby ameliorating the challenges posed by substantial disparities between the source and target datasets, commonly encountered in transfer learning scenarios. Empirical assessments, conducted through comparative experiments, substantiate the superior recognition capabilities of the proposed model compared to existing methodologies, underscoring its potential to enhance diagnostic accuracy in clinical settings. Despite the model's commendable performance, the issue of overfitting emerges due to the relatively limited size of the dataset. Nevertheless, this concern can be effectively mitigated through strategic measures such as dataset augmentation and meticulous fine-tuning of learning rates. These efforts are imperative for bolstering the model's generalization capabilities and fortifying its resilience against overfitting tendencies, thereby ensuring its applicability in real-world scenarios. Looking ahead, future research could delve into integrating computational features with semantic attributes to enable a comprehensive quantitative assessment of pulmonary nodules. This holistic approach holds promise for furnishing clinicians with robust diagnostic evidence, facilitating informed decision-making processes and potentially enhancing patient outcomes. By advancing the amalgamation of computational prowess with clinical semantics, this research trajectory heralds a paradigm shift towards more nuanced and efficacious diagnostic methodologies in the realm of pulmonary medicine.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

[1] Y. Yang, A. Gilliam, E. B. Ludmir, and K. Roberts, "Exploring the Generalization of Cancer Clinical Trial Eligibility Classifiers Across Diseases," arXiv preprint arXiv:2403.17135, 2024.

[2] Y. Zi, Q. Wang, Z. Gao, X. Cheng, and T. Mei, "Research on the Application of Deep Learning in Medical Image Segmentation and 3D Reconstruction," Academic Journal of Science and Technology, vol. 10, no. 2, pp. 8-12, 2024.

[3] Y. Yang, S. Jayaraj, E. Ludmir, and K. Roberts, "Text Classification of Cancer Clinical Trial Eligibility Criteria," in AMIA Annual Symposium Proceedings, vol. 2023, p. 1304, American Medical Informatics Association, 2023.

[4] S. Lu, Z. Liu, T. Liu, and W. Zhou, "Scaling-up medical vision-and-language representation learning with federated learning," Engineering Applications of Artificial Intelligence, vol. 126, p. 107037, 2023.

[5] K. Li, A. Zhu, W. Zhou, P. Zhao, J. Song, and J. Liu, "Utilizing Deep Learning to Optimize Software Development Processes," arXiv preprint arXiv:2404.13630, 2024.

[6] Y. Zhang, P. Ji, A. Wang, J. Mei, A. Kortylewski, and A. Yuille, "3d-aware neural body fitting for occlusion robust 3d human pose estimation," in Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 9399-9410, 2023.

[7] Z. Zhao, H. Yu, C. Lyu, P. Ji, X. Yang, and W. Yang, "Cross-Modal 2D-3D Localization with Single-Modal Query," in IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium, pp. 6171-6174, IEEE, July 2023.

[8] G. Lan, X. Y. Liu, Y. Zhang, and X. Wang, "Communication-efficient federated learning for resource-constrained edge devices," IEEE Transactions on Machine Learning in Communications and Networking, 2023.

[9] A. Zhu, K. Li, T. Wu, P. Zhao, W. Zhou, and B. Hong, "Cross-Task Multi-Branch Vision Transformer for Facial Expression and Mask Wearing Classification," arXiv preprint arXiv:2404.14606, 2024.

[10] J. Jin, H. Xu, P. Ji, and B. Leng, "IMC-NET: Learning Implicit Field with Corner Attention Network for 3D Shape Reconstruction," in 2022 IEEE International Conference on Image Processing (ICIP), pp. 1591-1595, IEEE, October 2022.

[11] T. Xu, H. Yang, and T. Wu, "Advancements in AI for Oncology: Developing an Enhanced YOLOv5-based Cancer Cell Detection System," International Journal of Innovative Research in Computer Science and Technology (IJIRCST), vol. 12, no. 2, pp. 75-80, 2024. [Online]. Available: doi:10.55524/ijircst.2024.12.2.13

[12] S. Dlamini, Y. H. Chen, and C. F. J. Kuo, "Complete fully automatic detection, segmentation and 3D reconstruction of tumor volume for non-small cell lung cancer using YOLOv4 and region-based active contour model," Expert Systems with Applications, vol. 212, p. 118661, 2023.

[13] M. Xiao, Y. Li, X. Yan, M. Gao, and W. Wang, "Convolutional neural network classification of cancer cytopathology images: taking breast cancer as an example," arXiv preprint arXiv:2404.08279, 2024.

[14] A. Zhu, J. Li, and C. Lu, "Pseudo view representation learning for monocular RGB-D human pose and shape estimation," IEEE Signal Processing Letters, vol. 29, pp. 712-716, 2021.

[15] X. Yan, W. Wang, M. Xiao, Y. Li, and M. Gao, "Survival Prediction Across Diverse Cancer Types Using Neural Networks," arXiv preprint arXiv:2404.08713, 2024.

[16] X. S. Wang and B. P. Mann, "Attractor Selection in Nonlinear Energy Harvesting Using Deep Reinforcement Learning," arXiv preprint arXiv:2010.01255, 2020.

[17] C. Szegedy et al., "Going deeper with convolutions," in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–9, IEEE, Boston, 2015. doi:10.1109/CVPR.2015.7298594

[18] C. Wang, D. Chen, L. Hao, X. Liu, Y. Zeng, J. Chen, and G. Zhang, "Pulmonary image classification based on inception-v3 transfer learning model," IEEE Access, vol. 7, pp. 146533-146541, 2019.

[19] X. S. Wang, J. D. Turner, and B. P. Mann, "Constrained attractor selection using deep reinforcement learning," Journal of Vibration and Control, vol. 27, no. 5-6, pp. 502-514, 2021.

[20] Q. Ning et al., "Rapid segmentation and sensitive analysis of CRP with paper-based microfluidic device using machine learning," Analytical and Bioanalytical Chemistry, vol. 414, no. 13, pp. 3959-3970, 2022.

[21] J. Yao, T. Wu, and X. Zhang, "Improving depth gradient continuity in transformers: A comparative study on monocular depth estimation with cnn," arXiv preprint arXiv:2308.08333, 2023.

[22] G. Lan, H. Wang, J. Anderson, C. Brinton, and V. Aggarwal, "Improved Communication Efficiency in Federated Natural Policy Gradient via ADMM-based Gradient Updates," Advances in Neural Information Processing Systems, vol. 36, 2024.

[23] S. G. Armato III et al., "Data From LIDC-IDRI," The Cancer Imaging Archive. [Online]. Available: https://doi.org/10.7937/K9/TCIA.2015.LO9QL9SX, 2015.
[24] Y. Li, X. Yan, M. Xiao, W. Wang, and F. Zhang, "Investigation of Creating Accessibility Linked Data Based on Publicly Available Accessibility Datasets," in Proceedings of the 2023 13th International Conference on Communication and Network Security, pp. 77–81, Association for Computing Machinery, 2024.
[25] M. Souto, P. G. Tahoces, S. C. JJ, M. J. Lado, M. Remy-Jardin, J. Remy, and J. J. Vidal, "Automatic detection of pulmonary nodules on computed tomography: a preliminary study," Radiologia, vol. 50, no. 5, pp. 387-392, 2008.
[26] C. H. Lee and J. S. Jwo, "Automatic segmentation for pulmonary nodules in CT images based on multifractal analysis," IET Image Processing, vol. 14, no. 7, pp. 1347-1353, 2020.