

RL2NdgsNet: Reinforcement learning based efficient classifier for mediastinal lymph nodes malignancy detection in CT images

Posina Praneeth
*Electronics and Communication
 Koneru Lakshmaiah Education
 Foundation
 Hyderabad, India
 190340065@klh.edu.in*

Dr. Shrish Verma
*Electronics and Communication
 National Institute of Technology
 Raipur, India
 shrishverma@nitrr.ac.in*

Abothula Dhana Lakshmi
*Electronics and Communication
 Koneru Lakshmaiah Education
 Foundation
 Hyderabad, India
 190349110@klh.edu.in*

Dr. Narendra D. Londhe
*Electrical Engineering
 National Institute of Technology
 Raipur, India
 nlondhe.ele@nitrr.ac.in*

Dr. Hitesh Tekchandani
*Electronics and Communication
 Koneru Lakshmaiah Education
 Foundation
 Hyderabad, India
 hitesh@klh.edu.in*

Abstract—Diagnosis of lymph nodes (LN)s in benign and malignant categories significantly impact the treatment planning of cancer. Invasive methods like biopsy are complex and painful procedures for cancer detection. Radiological modalities (like X-ray, USG, CT, and MRI) based medical imaging approaches are non-invasive and painless procedures for the cancer detection. Deep learning (DL) based architectures significantly improved the performance of medical imaging related tasks like organ detection and disease diagnosis. This paper introduces the reinforcement learning (RL) based DL network named – RL2NdgsNet (reinforcement learning based lymph nodes diagnosis network), for diagnosis of mediastinal lymph nodes (MLNs) in benign and malignant categories. Furthermore, we have designed custom DQN policy for the proposed RL2NdgsNet and also experimented with various state-of-the-art activation functions and exploration fraction. The proposed RL2NdgsNet achieves best results of sensitivity = 98.03%, specificity = 98.35%, accuracy = 98.2%, and AUC = 98.19%.

Keywords—Reinforcement learning, Lymph nodes, Deep learning, Cancer, CT.

I. INTRODUCTION

Lymph nodes (LN)s are significant part of human body's immune system. Malignancy detection in LNs is important step of cancer treatment planning. In current oncological practice diagnosis of LN in benign and malignant categories is done by invasive procedures like biopsy and fine needle aspiration cytology (FNAC). These invasive procedures having limitations of pain, risk associate with anesthesia, nerve or vessel injury, cost, time consuming, hospital stay, and complexity. Particularly detection of malignancy in mediastinal lymph nodes (MLNs) through invasive procedures is very complex due to their complicated anatomical location. Medical imaging (like X-ray, USG, CT, and MRI) based diagnosis approaches are non-invasive and painless procedures for the cancer detection. Computed tomography (CT) is easily available, cost effective and popularly used radiological modality for cancer detection. Hence, in this paper the authors have worked on MLNs malignancy detection in CT images. But manual inspection of medical images are time consuming, fatigue, and requires highly expert knowledge.

II. RELATED WORKS

To overcome these shortcomings, many researchers have developed image processing based algorithms for automated detection of malignancy in MLNs. These algorithms can be grouped into size based [1]–[5], CT texture feature based [6]–[9], and radiomics feature based [10]–[12] approaches.

The increased size can't be considered as the malignancy symptoms, because the increased size can be due to some viral infections also. Further, the measurement and analysis of size can be affected by inter/intra observer variability due to the prerequisite of expert anatomical knowledge, different guidelines, and calibration of different measuring tools. Additionally, approaches based on nodal size also suffer from very low sensitivity (41 to 67%), and specificity (65-94%).

Computed tomography texture analysis (CTTA) based classification rely on the quantification of tissue heterogeneity. In the small tumours their insufficient number of voxels available for useful heterogeneity analysis, hence the discriminative ability of the CTTA was compromised. Hence, CTTA methods can be concluded to have less promising results like accuracy (56% to 71%), sensitivity (52% to 81%), and specificity (60% to 97%) and also suffer from the trade-off between them.

Radiomics approaches bring up the quantitatively large-throughput feature space for extensive feature extraction and analysis. Despite having some potential, radiomics approaches have many short-comings such as dependency on reconstruction algorithms, artifacts, different acquisition modes, and feature selection bias. The sensitivity, specificity, and accuracy of radiomics approaches range from 68.01% to 94.8%, 73.35% to 92%, and 91.1%, respectively.

Deep learning (DL) is revolutionizing the complete technology industry and putting a comparable influence on medical imaging related tasks. The large computation and use of pre-trained weights from natural image datasets limits the novelty and doesn't justify the selection. Few researchers [13]–[17] applied DL for the same intended task of malignancy detection in MLNs. Above mentioned limitations of the existing literatures underscore the need of significant DL method for performance improvement. Reinforcement learning (RL) is the state-of-the-art DL approach that involves

learning of mapping situations to actions for maximizing the rewards. Based on these rewards next actions are decided. In this work, the authors have proposed reinforcement learning (RL) [18] based DL network named – RL2NdgsNet (reinforcement learning based lymph nodes diagnosis network), for diagnosis of mediastinal lymph nodes (MLNs) in benign and malignant categories.

III. METHODOLOGY

For the experimentations, the MLNs dataset is taken from [8]. As the DL approaches are data driven techniques and the size of the used dataset is small, hence to populate the dataset the authors have utilised generative adversarial network (GAN) [19] based data augmentation approach. The workflow of the proposed methodology is shown below in Fig. 1.

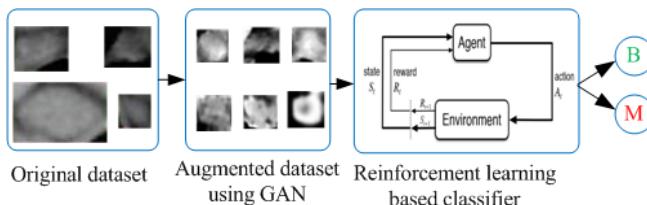


Fig. 1. Workflow of the proposed approach. “Benign, B”, “Malignant, M”.

RL algorithm popularly used for decision making. It is about taking suitable action to maximize reward in a user defined environment. Interactions with the environment and inspection of its responses are used to learn the optimal behaviour. In the proposed approach we are using Deep Q Networks (DQN) which helps in effective prediction of Q values for the proposed task. The Fig. 2 [20] shows the overview of the DQN based RL algorithm. The agent interacts with the Environment and observes current state S_t . An action a is chosen by deep neural network (DNN) which results in transition of state to S_{t+1} and receive the Reward r .

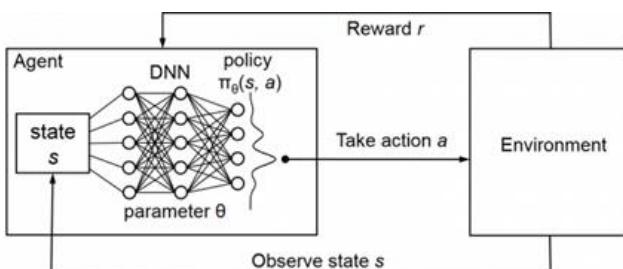


Fig. 2. Workflow of the DQN based RL algorithm.

The Q-Learning [18] algorithm involves learning a Q-table ('Q' is referred for quality) including the expected future reward for each state and action. The agent uses the Q-table to make the best decision for each state by selecting the action that maximises the expected future rewards. A Q-table is first randomly initialized. The number of columns represents the number of actions that can be taken, whereas the number of rows represents the number of states. An episode of the algorithm ends when the given state is a final or terminal state. If the episode isn't over, the agent in state S picks the action a that maximises the expected future rewards according to the Q-table. Sometimes, the agent is given the option of selecting an action at random implementing epsilon greedy strategy. It allows the agent to explore the environment. The agent adjusts the function $Q(S,a)$ by observing the outcome state S_{t+1} (if S_t

is not the terminal state) and reward r . This can be mathematically represented as following equation 1:

$$Q_{new}(S_t, a_t) = Q_{old}(S_t, a_t) + \alpha \cdot (r_t + \gamma \cdot \max Q(S_{t+1}, a) - Q_{old}(S_t, a_t)) \quad (1)$$

Where, α is the learning rate, r_t is the reward, γ is the discount factor, and $\max Q(S_{t+1}, a)$ is the highest estimated Q-value.

The limitation of Q learning is that it becomes inefficient for complicated environments with multiple possibilities and outcomes. Deep RL solves this problem by merging the Q learning approach with DL models. The basis of deep Q-learning [21] is the training of a neural network for estimating the various Q values for each action given a state. From this training the optimal Q values and network parameters will be learned.

The DQN architecture contains two neural networks, namely Q network and the Target (T) network, and also an Experience Replay component [21], [22]. The Q network is the agent that is trained to generate the optimal state-action value. The architecture of these networks is the same, but the weights are different. The weights from the Q network are replicated to the target network for every N steps. Using both these networks, the learning process becomes more stable, and the algorithm learns more effectively [18].

Experience Replay feature was introduced to make network updates more stable [21]–[23]. The transitions are added to a circular buffer called the replay buffer (training data) at each time step of data collection. Then, instead of computing the loss and gradient just using the most recent transition, we compute them using a mini-batch of transitions sampled from the replay buffer during training. This has two benefits, improved data efficiency by reusing each transition in several updates, and improved stability by using uncorrelated transitions in a batch.

Epsilon greedy strategy is used to achieve the balance between exploitation and exploration. We set an exploration rate (ϵ) initially [24]. This exploration rate represents the likelihood that our agent will explore rather than exploit the environment. At the start of each new episode, the agent learns more about the environment. ϵ will decay by some rate that we choose and likelihood of exploration drops. The agent becomes greedy in terms of exploiting it.

The Fig. 3 shows the workflow of the DQN [21], [22]. Experience replay chooses an ϵ -greedy action (benign or malignant) for the current image state, executes it in the environment, and returns a reward as well as the next image state.

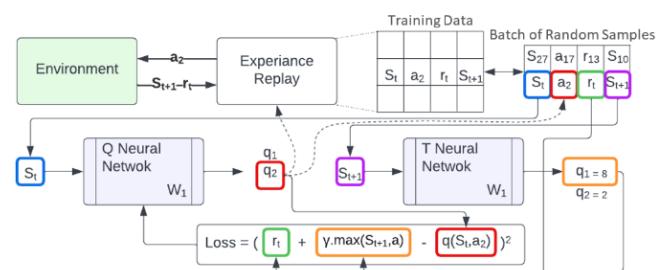


Fig. 3. Workflow of the DQN network.

This observation is saved as a sample of training data. To generate a training sample, the Q network operates as the agent while interacting with the environment. During this time, there is no DQN training. We now select a random batch of samples from this training data, ensuring that it includes both older and newer samples which helps in generalization of training data. The training data from this batch is then sent into both networks i.e., Q network and T network. The Q network predicts the Q value based on the current image state and action in each data sample. The Target network predicts the optimal Q value among all possible actions based on the next state in each data sample. Target Q value is the sum of optimal Q value for next state and the reward observed. To train the Q network, the loss is computed using the predicted Q Value and the target Q value. In this process, T network is not being trained. It is only after few specified time steps the Q network weights are replicated to the T network. This improves the T network's ability to predict optimal Q values. The loss function is determined using the following equation 2:

$$L = (r_t + \gamma \cdot Q(S_{t+1}, a) - Q(S_t, a_t))^2 \quad (2)$$

Where, L is Loss, $r_t + \gamma \cdot Q(S_{t+1}, a)$ is target Q value and $Q(S_t, a_t)$ is predicted Q value.

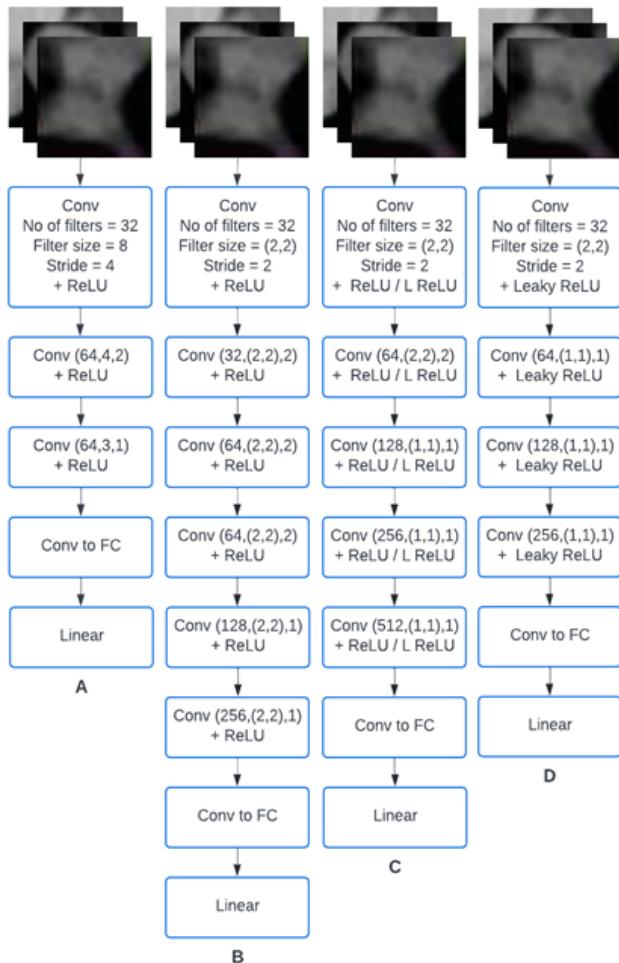


Fig. 4. Policy CNN Extractor Architectures. A) CnnPolicy / LnCnnPolicy Extractor B) Custom CNN Extractor 1 C) Custom CNN Extractor 2 D) Proposed CNN Extractor.

Q network is a neural network model that will be trained to predict Q values for all actions. This is accomplished through CNN Policy [25]. CNN policy is a DQN

implementation policy class. In this paper we are using the Stable baselines library [26] to get the CNN Policy for efficient training of the agent. In the selected policy, a `cnn_extractor` attribute for feature extraction of the input scaled images is present. Fig. 4(A) [25] shows the implemented `cnn_extractor` attribute architecture. LnCnnPolicy is another policy class which further supports layer normalisation.

To have more flexibility over the network architecture, in this paper we have proposed a custom CNN Policy. In this custom CNN policy a modified `cnn_extractor` architecture is utilised which requires scaled images as input tensor. This architecture will return a linear tensor from which features are extracted. The network architecture of our proposed custom CNN policy is depicted in Fig. 4(D). Fig. 4 shows the implemented and proposed policy CNN extractor architectures. Number of convolution layers, number of filters in each layer, filter size, stride value and activation function were modified several times to obtain an optimal architecture which resulted in better classification of the given task.

IV. EXPERIMENTAL SETUP AND DATASET DETAILS

The Gym [27] framework is utilised for evaluating the proposed RL2NdgsNet. The gym library consists the collection of various test problem environments. A gym environment is helpful in creating or developing the environment and defining actions.

The initial state of the RL problem is defined by describing the type and shape of our `action_space` [27]. We set this to Discrete - 2 since we have two classes of legitimate actions, benign and malignant. Similarly, observation space [27] has been defined, which contains all of the data in the environment that the agent will observe. Because we are dealing with the image it is set as `Box(0,255,[height,width,1])` for greyscale pixels. Later the agent's rewards are calculated, for a correct label prediction, the agent receives one point.

A. Dataset details

For the proposed study, MLNs dataset is taken from [8]. In this dataset there are total 271 MLNs CT images in which 138 are benign and 133 are malignant. As the DL approaches are data driven techniques and the size of the used dataset is small, hence to significantly populate the dataset, the GAN based augmentation approach was taken from our recent work [17]. (As the details of GAN based augmentation is properly mentioned in our recent work [17] hence it is not described again in this paper). After augmentation there are total 6000 images available for experimentation in which 3000 are benign and 3000 are malignant. The size of images of base dataset varies from 16x24 to 122x112 hence before feeding to GAN networks they are resized to 50x50.

V. RESULTS AND DISCUSSION

All the simulations were performed on the Google Colab platform using Keras [28] DL and Stable-baselines [26] RL library. Sensitivity (SEN), specificity (SPE), accuracy (ACC), and area under the curve (AUC) are calculated as the evaluation parameters. The results of all the proposed and implemented models are summarized in Table 1. The proposed CNN extractor architecture approach achieves ACC = 98.2%, SEN = 98.03%, SPE = 98.35% and AUC = 98.19%. It's observable from Table 1 that, the proposed methodology achieves improvement of 5.69% in ACC, 4.78% in SEN,

7.25% in SPE, and 6.01% in AUC compared to basic CnnPolicy.

Table 2 represents the performance metrics obtained from related works. It's noticeable from Table 2 that the proposed approach attains improvement of 3.20% in ACC, 4.03% in SEN, 1.35% in SPE, and 3.19% in AUC. Fig. 5 shows the time steps (represents 5400 training images) versus reward and loss curve while performing the training for Proposed CNN Extractor + Leaky ReLu + Lnorm ($\epsilon=0.8$) with respect to 5-fold validation. The first 1000 steps of the model are used to generate samples for the replay buffer. Upon reaching 1000 steps, learning begins. Hence, we find the curves starting from the 1k mark. The model gradually starts learning from there. The spikes at regular intervals in the loss curve are due to the replication of weights from the Q network to the T network for every 500 steps. The continuous increase in reward and decrease in loss values suggest the increase in performance by the model.

TABLE I. PERFORMANCE EVALUATION PARAMETERS OF ALL THE PROPOSED DQN POLICY MODELS.

Implemented DQN Policy (CNN Extractor Architecture)	Exploration fraction	ACC	SEN	SPE	AUC
		(In %)			
CnnPolicy + ReLU	0.1	92.51	93.25	91.1	92.18
LnCnnPolicy + ReLU + Layer Normalisation (LNorm)	0.1	93.3	93.96	92.7	93.34
LnCnnPolicy + ReLU + LNorm	0.8	94.08	94.13	93.93	94.03
Custom CNN Extractor 1 + ReLU + LNorm	0.8	95.33	95.14	95.58	95.36
Custom CNN Extractor 2 + ReLU + LNorm	0.8	96.5	96.44	96.58	96.51
Custom CNN Extractor 2 + Leaky ReLU + LNorm	0.8	97.46	97.94	96.99	97.46
Proposed CNN Extractor + Leaky ReLU + LNorm	1	97.50	97.98	97.04	97.51
Proposed CNN Extractor + Leaky ReLU + LNorm	0.8	98.20	98.03	98.35	98.19

TABLE II. PERFORMANCE METRICS OBTAINED FROM RELATED WORKS.

Authors	ACC	SEN	SPE	AUC
	(in %)			
Michael et al. [7]	-	53	97	83
Hamid [6]	71	81	80	87
Hongkai et al. [14]	86	84	88	91
Pham et al. [8]	70	75	90	89
Pham et al. [11]	-	68.01	73.35	75
Pham et al. (raw images) [15]	87.07	90.42	83.96	-
Hitesh et al. [13]	63.14	71.03	55.69	63
Hitesh et al. [16]	90	91	90	90
Hitesh et al. [17]	95	94	97	95
Proposed Model	98.20	98.03	98.35	98.19

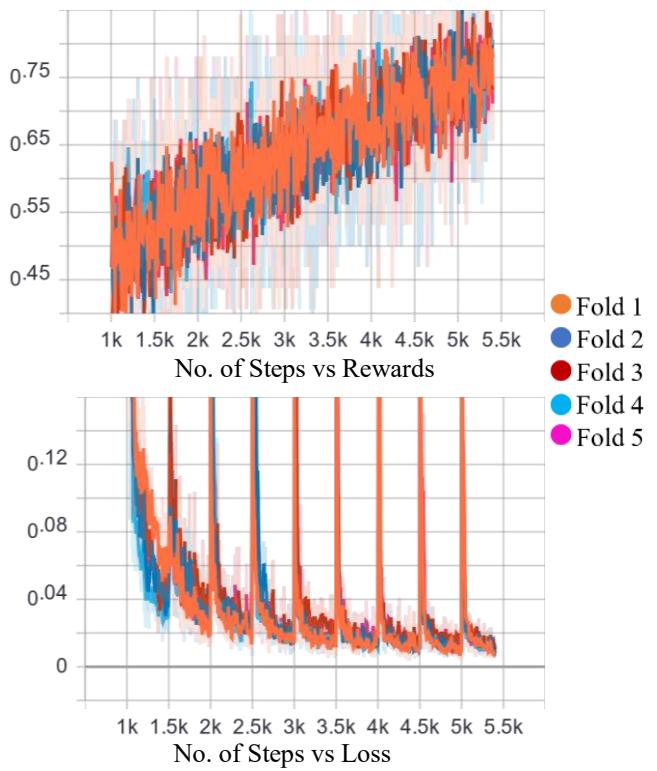


Fig. 5. The number of steps versus reward and loss curve of Proposed CNN Extractor + Leaky ReLu + Lnorm ($\epsilon=0.8$).

VI. CONCLUSION AND FUTURE SCOPE

In this paper the RL based DL architecture named RL2NdgsNet is proposed for classification of benign and malignant MLNs in CT images. The achieved results signifies the utility of the proposed modification in policy network provides. The utilised GAN based [17] augmentation approach solves the data scarcity problem. The wise selection of augmentation methods and their amalgamation surely leads to deeper and successful learning of networks. The optimized selection of activation function and exploration fraction improves the performance of proposed RL2NdgsNet. The results comparison with state-of-the-art related works validates the significance of the proposed approach. The proposed work limited to severity assessment of LN only, but may be nearby organs are also at risk. Hence, in future work, generalizable RL will be explored for more utility and applications of RL in healthcare.

REFERENCES

- [1] J. M. Seely, J. R. Mayo, R. R. Miller, and N. L. Müller, "T1 lung cancer: prevalence of mediastinal nodal metastases and diagnostic accuracy of CT," *Radiology*, vol. 186, no. 1, pp. 129–132, Jan. 1993, doi: 10.1148/radiology.186.1.8416552.
- [2] H. Libshitz and R. McKenna, "Mediastinal lymph node size in lung cancer," *American Journal of Roentgenology*, vol. 143, no. 4, pp. 715–718, Oct. 1984, doi: 10.2214/ajr.143.4.715.
- [3] H. C. Steinert et al., "Non-small cell lung cancer: nodal staging with FDG PET versus CT with correlative lymph node mapping and sampling," *Radiology*, vol. 202, no. 2, pp. 441–446, Feb. 1997, doi: 10.1148/radiology.202.2.9015071.
- [4] D. W. von Haag, D. M. Follette, P. F. Roberts, D. Shelton, L. D. Segel, and T. M. Taylor, "Advantages of Positron Emission Tomography over Computed Tomography in Mediastinal Staging of Non-Small Cell Lung Cancer," *Journal of Surgical Research*, vol. 103, no. 2, pp. 160–164, Apr. 2002, doi: 10.1006/jssr.2002.6354.
- [5] L. K. Toney and H. J. Vesselle, "Neural Networks for Nodal Staging of Non-Small Cell Lung Cancer with FDG PET and CT: Importance of Combining Uptake Values and Sizes of Nodes and Primary Tumor,"

- Radiology*, vol. 270, no. 1, pp. 91–98, Jan. 2014, doi: 10.1148/radiol.13122427.
- [6] H. Bayanati *et al.*, “Quantitative CT texture and shape analysis: Can it differentiate benign and malignant mediastinal lymph nodes in patients with primary lung cancer?,” *European Radiology*, vol. 25, no. 2, pp. 480–487, 2014, doi: 10.1007/s00330-014-3420-6.
- [7] M. B. Andersen, S. W. Harders, B. Ganeshan, J. Thygesen, H. H. T. Madsen, and F. Rasmussen, “CT texture analysis can help differentiate between malignant and benign lymph nodes in the mediastinum in patients suspected for lung cancer,” *Acta Radiologica*, vol. 57, no. 6, pp. 669–676, 2016, doi: 10.1177/0284185115598808.
- [8] T. D. Pham, Y. Watanabe, M. Higuchi, and H. Suzuki, “Texture Analysis and Synthesis of Malignant and Benign Mediastinal Lymph Nodes in Patients with Lung Cancer on Computed Tomography,” *Scientific Reports*, vol. 7, no. June 2016, pp. 1–10, 2017, doi: 10.1038/srep43209.
- [9] X. Gao *et al.*, “The method and efficacy of support vector machine classifiers based on texture features and multi-resolution histogram from 18F-FDG PET-CT images for the evaluation of mediastinal lymph nodes in patients with lung cancer,” *European Journal of Radiology*, vol. 84, no. 2, pp. 312–317, Feb. 2015, doi: 10.1016/j.ejrad.2014.11.006.
- [10] X. Yang *et al.*, “A new approach to predict lymph node metastasis in solid lung adenocarcinoma: a radiomics nomogram,” *J Thorac Dis*, vol. 10, no. Suppl 7, pp. S807–S819, Apr. 2018, doi: 10.21037/jtd.2018.03.126.
- [11] T. D. Pham, “Complementary features for radiomic analysis of malignant and benign mediastinal lymph nodes,” in *Proceedings - International Conference on Image Processing, ICIP*, 2018, vol. 2017-Septe, no. 4, pp. 3849–3853. doi: 10.1109/ICIP.2017.8297003.
- [12] Y. Zhong, M. Yuan, T. Zhang, Y.-D. Zhang, H. Li, and T.-F. Yu, “Radiomics Approach to Prediction of Occult Mediastinal Lymph Node Metastasis of Lung Adenocarcinoma,” *American Journal of Roentgenology*, vol. 211, no. 1, pp. 109–113, Jul. 2018, doi: 10.2214/AJR.17.19074.
- [13] H. Tekchandani, S. Verma, and N. D. Londhe, “Severity Assessment of Lymph Nodes in CT Images using Deep Learning Paradigm,” in *International Conference on Computing Methodologies and Communication*, 2018, pp. 686–691.
- [14] H. Wang *et al.*, “Comparison of machine learning methods for classifying mediastinal lymph node metastasis of non-small cell lung cancer from 18 F-FDG PET/CT images,” *EJNMMI research*, vol. 7, no. 1, p. 11, 2017.
- [15] T. D. Pham, “From Raw Pixels to Recurrence Image for Deep Learning of Benign and Malignant Mediastinal Lymph Nodes on Computed Tomography,” *IEEE Access*, 2021, doi: 10.1109/ACCESS.2021.3094577.
- [16] H. Tekchandani, S. Verma, and N. D. Londhe, “Mediastinal lymph node malignancy detection in computed tomography images using fully convolutional network,” *Biocybernetics and Biomedical Engineering*, vol. 40, no. 1, pp. 187–199, Jan. 2020, doi: 10.1016/j.bbe.2019.05.002.
- [17] H. Tekchandani, S. Verma, and N. Londhe, “Performance improvement of mediastinal lymph node severity detection using GAN and Inception network,” *Computer Methods and Programs in Biomedicine*, vol. 194, p. 105478, Oct. 2020, doi: 10.1016/j.cmpb.2020.105478.
- [18] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, Second. The MIT Press, 2014.
- [19] I. Goodfellow *et al.*, “Generative adversarial nets,” in *Advances in neural information processing systems*, 2014, pp. 2672–2680.
- [20] H. Mao, M. Alizadeh, I. Menache, and S. Kandula, “Resource Management with Deep Reinforcement Learning,” in *Proceedings of the 15th ACM Workshop on Hot Topics in Networks*, New York, NY, USA, Nov. 2016, pp. 50–56. doi: 10.1145/3005745.3005750.
- [21] V. Mnih *et al.*, “Human-level control through deep reinforcement learning,” *Nature*, vol. 518, no. 7540, Art. no. 7540, Feb. 2015, doi: 10.1038/nature14236.
- [22] V. Mnih *et al.*, “Playing Atari with Deep Reinforcement Learning,” *arXiv:1312.5602 [cs]*, Dec. 2013, Accessed: Apr. 29, 2022. [Online]. Available: <http://arxiv.org/abs/1312.5602>
- [23] T. Lahire, M. Geist, and E. Rachelson, “Large Batch Experience Replay,” *arXiv:2110.01528 [cs]*, Oct. 2021, Accessed: Apr. 29, 2022. [Online]. Available: <http://arxiv.org/abs/2110.01528>
- [24] S. Amin, M. Gomrokchi, H. Satija, H. van Hoof, and D. Precup, “A Survey of Exploration Methods in Reinforcement Learning,” *arXiv:2109.00157 [cs]*, Sep. 2021, Accessed: Apr. 28, 2022. [Online]. Available: <http://arxiv.org/abs/2109.00157>
- [25] “DQN — Stable Baselines 2.10.2 documentation.” https://stable-baselines.readthedocs.io/en/master/modules/dqn.html#stable_baselines.dqn.CnnPolicy (accessed Apr. 28, 2022).
- [26] “Welcome to Stable Baselines docs! - RL Baselines Made Easy — Stable Baselines 2.10.2 documentation.” <https://stable-baselines.readthedocs.io/en/master/index.html> (accessed Apr. 18, 2022).
- [27] G. Brockman *et al.*, “OpenAI Gym,” *arXiv:1606.01540 [cs]*, Jun. 2016, Accessed: Apr. 28, 2022. [Online]. Available: <http://arxiv.org/abs/1606.01540>
- [28] F. Chollet and others, “Keras: The Python Deep Learning library,” *Astrophysics Source Code Library*, p. ascl:1806.022, Jun. 2018.