



CARL: Cross-Aligned Representation Learning for Multi-view Lung Cancer Histology Classification

Yin Luo¹, Wei Liu¹, Tao Fang¹, Qilong Song², Xuhong Min², Minghui Wang¹(✉),
and Ao Li¹(✉)

¹ School of Information Science and Technology, University of Science and
Technology of China, Hefei 230026, China
{mhwang, aoli}@ustc.edu.cn

² Department of Radiology, Anhui Chest Hospital, Hefei 230039, Anhui, China

Abstract. Accurately classifying the histological subtype of non-small cell lung cancer (NSCLC) using computed tomography (CT) images is critical for clinicians in determining the best treatment options for patients. Although recent advances in multi-view approaches have shown promising results, discrepancies between CT images from different views introduce various representations in the feature space, hindering the effective integration of multiple views and thus impeding classification performance. To solve this problem, we propose a novel method called cross-aligned representation learning (CARL) to learn both view-invariant and view-specific representations for more accurate NSCLC histological subtype classification. Specifically, we introduce a cross-view representation alignment learning network which learns effective view-invariant representations in a common subspace to reduce multi-view discrepancies in a discriminability-enforcing way. Additionally, CARL learns view-specific representations as a complement to provide a holistic and disentangled perspective of the multi-view CT images. Experimental results demonstrate that CARL can effectively reduce the multi-view discrepancies and outperform other state-of-the-art NSCLC histological subtype classification methods.

Keywords: Cross-view Alignment · Representation Learning · Multi-view ·
Histologic Subtype Classification · Non-small Cell Lung Cancer

1 Introduction

Lung cancer is currently the foremost cause of cancer-related mortalities globally, with non-small cell lung cancer (NSCLC) being responsible for 85% of reported cases [25]. Within NSCLC, squamous cell carcinoma (SCC) and adenocarcinoma (ADC) are recognized as the two principal histological subtypes. Since SCC and ADC differ in the effectiveness of chemotherapy and the risk of complications, accurate identification of different subtypes is crucial for clinical treatment options [15]. Although pathological diagnosis via lung biopsy can provide a reliable result of subtype identification, it is

highly invasive with potential clinical implications [19]. Therefore, non-invasive methods utilizing computed tomography (CT) images have garnered significant attention over the last decade [15, 16].

Recently, several deep-learning methods have been put forward to differentiate between the NSCLC histological subtypes using CT images [4, 11, 13, 22]. Chaunzwa et al. [4] and Marentakis et al. [13] both employ a convolutional neural network (CNN) model with axial view CT images to classify the tumor histology into SCC and ADC. Albeit the good performance, the above 2D CNN-based models only take CT images from a single view as the input, limiting their ability to describe rich spatial properties of CT volumes [20]. Multi-view deep learning, a 2.5D method, represents a promising solution to this issue, as it focuses on obtaining a unified joint representation from different views of lung nodules to capture abundant spatial information [16, 20]. For example, Wu et al. [22] aggregate features from axial, coronal, and sagittal view CT images via a multi-view fusion model. Similarly, Li et al. [11] also extract patches from three orthogonal views of a lung nodule and present a multi-view ResNet for feature fusion and classification. By integrating multi-view representations, these methods efficiently preserve the spatial information of CT volumes while significantly reducing the required computational resource compared to 3D CNNs [9, 20, 23].

Despite the promising results of previous multi-view methods, they still confront a severe challenge for accurate NSCLC histological subtype prediction. In fact, due to the limitation of scan time and hardware capacity in clinical practice, different views of CT volumes are anisotropic in terms of in-plane and inter-plane resolution [21]. Additionally, images from certain views may inevitably contain some unique background information, e.g., the spine in the sagittal view [17]. Such anisotropy and background dissimilarity both reveal the existence of significant variations between different views, which lead to markedly various representations in feature space. Consequently, the discrepancies of distinct views will hamper the fusion of multi-view information, limiting further improvements in the classification performance.

To overcome the challenge mentioned above, we propose a novel cross-aligned representation learning (CARL) method for the multi-view histologic subtype classification of NSCLC. CARL offers a holistic and disentangled perspective of multi-view CT images by generating both view-invariant and -specific representations. Specifically, CARL incorporates a cross-view representation alignment learning network which targets the reduction of multi-view discrepancies by obtaining discriminative view-invariant representations. A shared encoder with a novel discriminability-enforcing similarity constraint is utilized to map all representations learned from multi-view CT images to a common subspace, enabling cross-view representation alignment. Such aligned projections help to capture view-invariant features of cross-view CT images and meanwhile make full use of the discriminative information obtained from each view. Additionally, CARL learns view-specific representations as well which complement the view-invariant ones, providing a comprehensive picture of the CT volume data for histological subtype prediction. We validate our approach by using a publicly available NSCLC dataset from The Cancer Imaging Archive (TCIA). Detailed experimental results demonstrate the effectiveness of CARL in reducing multi-view discrepancies and improving NSCLC

histological subtype classification performance. Our contributions can be summarized as follows:

- A novel cross-aligned representation learning method called CARL is proposed for NSCLC histological subtype classification. To reduce the discrepancies of multi-view CT images, CARL incorporates a cross-view representation alignment learning network for discriminative view-invariant representations.
- We employ a view-specific representation learning network to learn view-specific representations as a complement to the view-invariant representations.
- We conduct experiments on a publicly available dataset and achieve superior performance compared to the most advanced methods currently available.

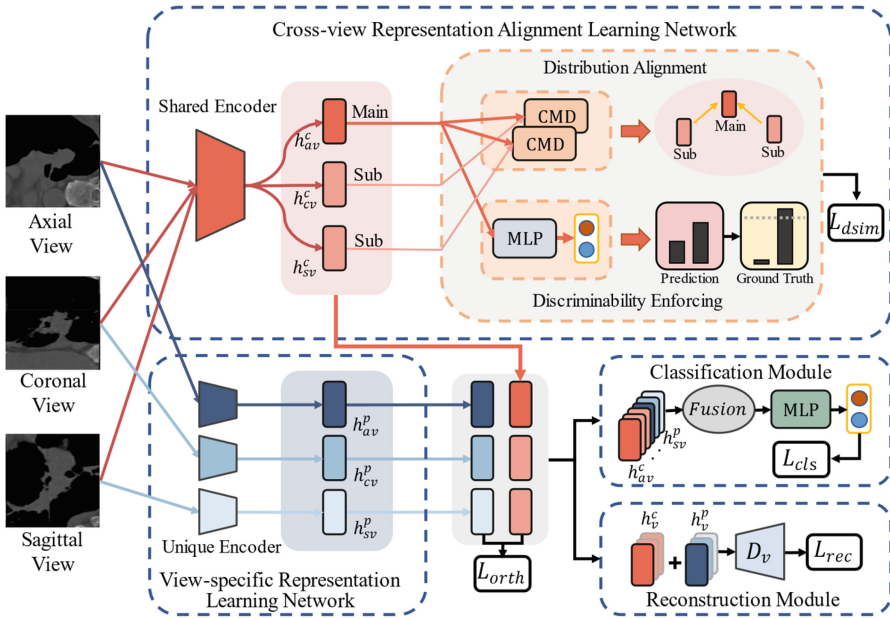


Fig. 1. Illustration of the proposed CARL.

2 Methodology

2.1 Architecture Overview

Figure 1 shows the overall architecture of CARL. The cross-view representation alignment learning network includes a shared encoder which projects patches of axial, coronal, and sagittal views into a common subspace with a discriminability-enforcing similarity constraint to obtain discriminative view-invariant representations for multi-view discrepancy reduction. In addition, CARL introduces a view-specific representation learning network consisting of three unique encoders which focus on learning view-specific

representations in respective private subspaces to yield complementary information to view-invariant representations. Finally, we introduce a histological subtype classification module to fuse the view-invariant and -specific representations and make accurate NSCLC histological subtype classification.

2.2 Cross-View Representation Alignment Learning

Since the discrepancies of different views may result in divergent statistical properties in feature space, e.g., huge distributional disparities, aligning representations of different views is essential for multi-view fusion. With the aim to reduce multi-view discrepancies, CARL introduces a cross-view representation alignment learning network for mapping the representations from distinct views into a common subspace, where view-invariant representations can be obtained by cross-view alignment. Specifically, inspired by [12, 14, 24], we exert a discriminability-enforcing similarity constraint to align all sub-view representations with those of the main view, significantly mitigating the distributional disparities of multi-view representations.

Technically speaking, given the axial view image I_{av} , coronal view image I_{cv} , and sagittal view image I_{sv} , the cross-view representation alignment learning network tries to generate view-invariant representations h_v^c , $v \in \{av, cv, sv\}$ via a shared encoder based on a residual neural network [10]. This can be formulated as below:

$$h_v^c = E_c(I_v), \quad v \in \{av, cv, sv\} \quad (1)$$

where $E_c(\cdot)$ indicates the shared encoder, and av, cv, sv represent axial, coronal, and sagittal views, respectively. In the common subspace, we hope that through optimizing the shared encoder $E_c(\cdot)$, the view-invariant representations can be matched to some extent. However, the distributions of h_{av}^c , h_{cv}^c and h_{sv}^c are very complex due to the significant variations between different views, which puts a burden on obtaining well-aligned view-invariant representations with merely an encoder.

To address this issue, we design a discriminability-enforcing similarity loss L_{sim} to further enhance the alignment of cross-view representations in the common subspace. Importantly, considering that the axial view has a higher resolution than other views and are commonly used in clinical diagnosis, we choose axial view as the main view and force the sub-views (e.g., the coronal and sagittal views) to seek distributional similarity with the main view. Mathematically, we introduce a cross-view similarity loss L_{sim} which calculates the central moment discrepancy (CMD) metric [24] between all sub-views and the main view as shown below:

$$L_{sim} = \frac{1}{N} \sum_{i=1}^N CMD(h_i^{sub}, h^{main}) \quad (2)$$

where $CMD(\cdot)$ denotes the distance metric which measures the distribution disparities between the representations of i -th sub-view h_i^{sub} and the main view h^{main} . N is the number of sub-views. Despite the fact that minimizing the L_{sim} can efficiently mitigate the issue of distributional disparities, it may not guarantee that the alignment network will learn informative and discriminative representations. Inspired by recent work on

multimodal feature extraction [12, 14], we impose a direct supervision by inputting h^{main} into a classifier $f(\cdot)$ to obtain the prediction of histological subtype, and use a cross-entropy loss to enforce the discriminability of the main-view representations. Finally, the discriminability-enforcing similarity loss L_{dsim} is as follows:

$$L_{dsim} = L_{sim} + \lambda \cdot L_{CE}\left(f\left(h^{main}\right), y\right) \quad (3)$$

where y denotes the ground-truth subtype labels, λ controls the weight of L_{CE} . We observed that L_{CE} is hundred times smaller than L_{sim} , so this study uses an empirical value of $\lambda = 110$ to balance the magnitude of two terms. By minimizing L_{dsim} , the cross-view representation alignment learning network pushes the representations of each sub-view to align with those of the main view in a discriminability-enforcing manner. Notably, the benefits of such cross-alignment are twofold. Firstly, it greatly reduces the discrepancies between the sub-views and the main view, leading to consistent view-invariant representations. Secondly, since the alignment between distinct views compels the representation distribution of the sub-views to match that of the discriminative main view, it can also enhance the discriminative power of the sub-view representations. In other words, the cross-alignment procedure spontaneously promotes the transfer of discriminative information learned by the representations of the main view to those of the sub-views. As a result, the introduced cross-view representation alignment learning network is able to generate consistent and discriminative view-invariant representations cross all views to effectively narrow the multi-view discrepancies.

2.3 View-Specific Representation Learning

On the basis of learning view-invariant representations, CARL additionally learns view-specific representations in respective private subspaces, which provides supplementary information for the view-invariant representations and contribute to subtype classification as well. To be specific, a view-specific representation learning network containing three unique encoders is proposed to learn view-specific representations $h_v^p, v \in \{av, cv, sv\}$, enabling effective exploitation of the specific information from each view. We formulate the unique encoders as follows:

$$h_v^p = E_v^p(I_v), v \in \{av, cv, sv\} \quad (4)$$

where $E_v^p(\cdot)$ is the encoder function dedicated to capture single-view characteristics.

To induce the view-invariant and -specific representations to learn unique characteristics of each view, we draw inspiration from [14] and adopt an orthogonality loss L_{orth} with the squared Frobenius norm between the representations in the common and private subspaces of each view, which is denoted by $L_{orth} = \|h_v^c h_v^p\|_F^2, v \in \{av, cv, sv\}$. A reconstruction module is also employed to calculate a reconstruction loss L_{rec} between original image I_v and reconstructed image I_v^r using the L_1 -norm, which ensures the hidden representations to capture details of the respective view.

2.4 Histologic Subtype Classification

After obtaining view-invariant and -specific representations from each view, we integrate them together to perform NSCLC subtype classification. Specifically, we apply a residual

block [10] to fuse view-invariant and -specific representations into a unified multi-view representation h . Then, h is sent to a multilayer perceptron neural network (MLP) to make the precise NSCLC subtype prediction. The NSCLC histological subtype classification loss L_{cls} can be calculated by using cross-entropy loss.

2.5 Network Optimization

The optimization of CARL is achieved through a linear combination of several loss terms, including discriminability-enforcing similarity loss L_{dsim} , orthogonality loss L_{orth} , reconstruction loss L_{rec} and the classification loss L_{cls} . Accordingly, the total loss function can be formulated as a weighted sum of these separate loss terms:

$$L_{total} = L_{cls} + \alpha L_{dsim} + \beta L_{orth} + \gamma L_{rec} \quad (5)$$

where α , β and γ denote the weights of L_{dsim} , L_{rec} and L_{orth} . To normalize the scale of L_{dsim} which is much larger than the other terms, we introduce a scaling factor $S = 0.001$, and perform a grid search for α , β and γ in the range of 0.1S-S, 0.1-1, and 0.1-1, respectively. Throughout the experiments, we set the values of α , β and γ to 0.6S, 0.4 and 0.6, respectively.

3 Experiments and Results

3.1 Dataset

Our dataset NSCLC-TCIA for lung cancer histological subtype classification is sourced from two online resources of The Cancer Imaging Archive (TCIA) [5]: NSCLC Radiomics [1] and NSCLC Radiogenomics [2]. Exclusion criteria involves patients diagnosed with large cell carcinoma or not otherwise specified, along with cases that have contouring inaccuracies or lacked tumor delineation [9, 13]. Finally, a total of 325 available cases (146 ADC cases and 179 SCC cases) are used for our study. We evaluate the performance of NSCLC classification in five-fold cross validation on the NSCLC-TCIA dataset, and measure accuracy (Acc), sensitivity (Sen), specificity (Spe), and the area under the receiver operating characteristic (ROC) curve (AUC) as evaluation metrics. We also conduct analysis including standard deviations and 95% CI, and DeLong statistical test for further AUC comparison.

For preprocessing, given that the CT data from NSCLC-TCIA has an in-plane resolution of $1 \text{ mm} \times 1 \text{ mm}$ and a slice thickness of 0.7–3.0 mm, we resample the CT images using trilinear interpolation to a common resolution of $1 \text{ mm} \times 1 \text{ mm} \times 1 \text{ mm}$. Then one 128×128 pixel slice is cropped from each view as input based on the center of the tumor. Finally following [7], we clip the intensities of the input patches to the interval $(-1000, 400 \text{ Hounsfield Unit})$ and normalize them to the range of $[0, 1]$.

3.2 Implementation Details

The implementation of CARL is carried out using PyTorch and run on a workstation equipped with Nvidia GeForce RTX 2080Ti GPUs and Intel Xeon CPU 4110 @ 2.10GHz. Adam optimizer is used with an initial learning rate of 0.00002, and the batch size is set to 8.

3.3 Results

Comparison with Existing Methods. Several subtype classification methods have been employed for comparison including: two conventional methods, four single-view and 3D deep learning methods, and four representative multi-view methods. We use publicly available codes of these comparison methods and implement models for methods without code. The experimental results are reported in Table 1. The multi-view methods are generally superior to the single-view and 3D deep learning methods. It illustrates that the multi-view methods can exploit richer spatial properties of CT volumes than the single-view methods while greatly reducing the model parameters to avoid overfitting compared to the 3D methods. The floating point operations (FLOPs) comparison between CARL (0.9 GFLOPs) and the 3D method [9] (48.4 GFLOPs) also proves the computational efficiency of our multi-view method. Among all multi-view methods, our proposed CARL achieves the best results, outperforming Wu et al. by 3.2%, 3.2%, 1.5% and 4.1% in terms of AUC, Acc, Sen and Spe, respectively. Not surprisingly, the ROC curve of CARL in Fig. 2(a) is also closer to the upper-left corner, further indicating its superior performance. These results demonstrate that CARL can effectively narrow the discrepancies of different views by obtaining view-invariant representations in a discriminative way, thus leading to excellent classification accuracy compared to other methods.

Table 1. Results on NSCLC-TCIA dataset when CARL was compared with other SOTA methods using five-fold cross validation. * indicates the p-value is less than 0.05 in DeLong test between the AUC of compared method and CARL.

Category	Methods	AUC	95% CI	Acc	Sen	Spe
Conventional	RF [3]	0.742 ± 0.061*	0.684–0.791	0.667	0.623	0.703
	SVM [6]	0.756 ± 0.069*	0.714–0.817	0.699	0.670	0.725
Deep learning	Chaunzwa et al. [4]	0.774 ± 0.051*	0.710–0.816	0.713	0.692	0.729
	Marentakis et al. [13]	0.770 ± 0.076*	0.707–0.813	0.715	0.663	0.757
	Yanagawa et al. [23]	0.777 ± 0.062*	0.718–0.822	0.719	0.650	0.776
	Guo et al. [9]	0.772 ± 0.072*	0.676–0.783	0.676	0.635	0.712
	MVCNN [18]	0.784 ± 0.052*	0.707–0.811	0.691	0.637	0.733
Multi-view	GVCNN [8]	0.767 ± 0.054*	0.704–0.809	0.685	0.581	0.770
	Wu et al. [22]	0.785 ± 0.080	0.744–0.844	0.736	0.717	0.756
	Li et al. [11]	0.782 ± 0.069*	0.719–0.822	0.729	0.705	0.748
	CARL	0.817 ± 0.055	0.770–0.862	0.768	0.732	0.797

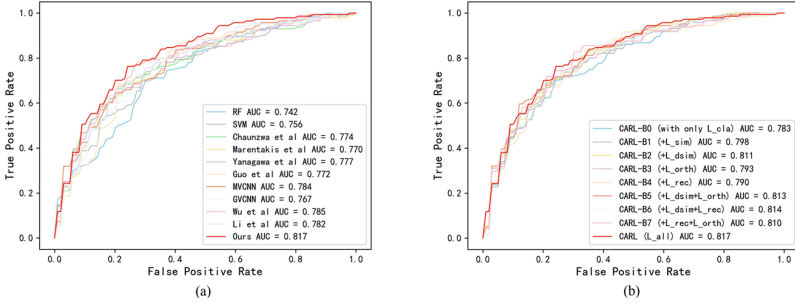


Fig. 2. ROC plots of (a) compared methods and (b) ablation analysis on NSCLC-TCIA dataset.

Table 2. Results of ablation analysis on the NSCLC-TCIA dataset.

Methods	AUC	Acc	Sen	Spe
CARL-B0 (with only L_{cla})	0.783	0.713	0.670	0.748
CARL-B1 ($+L_{sim}$)	0.798	0.734	0.726	0.741
CARL-B2 ($+L_{dsim}$)	0.811	0.738	0.705	0.764
CARL-B3 ($+L_{orth}$)	0.793	0.722	0.705	0.736
CARL-B4 ($+L_{rec}$)	0.790	0.701	0.697	0.705
CARL-B5 ($+L_{dsim}+L_{orth}$)	0.813	0.756	0.725	0.782
CARL-B6 ($+L_{dsim}+L_{rec}$)	0.814	0.747	0.732	0.758
CARL-B7 ($+L_{orth}+L_{rec}$)	0.810	0.752	0.712	0.786
CARL (L_{all})	0.817	0.768	0.732	0.797

Ablation Analysis. We evaluate the efficacy of different losses in our method. The results are reported in Table 2, where CARL-B0 refers to CARL only using the classification loss, $+L_*$ indicates the loss superimposed on CARL-B0 and L_{all} denotes that we utilize all the losses in Eq. 5. We can observe that CARL-B2 performs better than CARL-B0 by employing the discriminability-enforcing similarity loss to align cross-view representations. Besides, CARL-B3 and CARL-B4 show better performance than CARL-B0, illustrating view-specific representations as a complement which can also contribute to subtype classification. Though single loss already contributes to performance improvement, CARL-B5 to CARL-B7 demonstrate that the combinations of different losses can further enhance classification results. More importantly, CARL with all losses achieves the best performance among all methods, demonstrating that our proposed method effectively reduces multi-view discrepancies and significantly improves the performance of histological subtype classification by providing a holistic and disentangled perspective of the multi-view CT images. The ROC curve of CARL in Fig. 2(b) is generally above its variants, which is also consistent with the quantitative results.

4 Conclusion

In summary, we propose a novel multi-view method called cross-aligned representation learning (CARL) for accurately distinguishing between ADC and SCC using multi-view CT images of NSCLC patients. It is designed with a cross-view representation alignment learning network which effectively generates discriminative view-invariant representations in the common subspace to reduce the discrepancies among multi-view images. In addition, we leverage a view-specific representation learning network to acquire view-specific representations as a necessary complement. The generated view-invariant and -specific representations together offer a holistic and disentangled perspective of the multi-view CT images for histological subtype classification of NSCLC. The experimental result on NSCLC-TCIA demonstrates that CARL reaches 0.817 AUC, 76.8% Acc, 73.2% Sen, and 79.7% Spe and surpasses other relative approaches, confirming the effectiveness of the proposed CARL method.

Acknowledgement. This work was supported in part by the National Natural Science Foundation of China under Grants 61971393, 62272325, 61871361 and 61571414.

References

1. Aerts, H.J.W.L. et al.: Decoding tumour phenotype by noninvasive imaging using a quantitative Radiomics approach. *Nat. Commun.* **5**(1), 4006 (2014). <https://doi.org/10.1038/ncomms5006>
2. Bakr, S. et al.: A radiogenomic dataset of non-small cell lung cancer. *Sci. Data* **5**(1), 180202 (2018). <https://doi.org/10.1038/sdata.2018.202>
3. Breiman, L.: Random forests. *Mach. Learn.* **45**(1), 5–32 (2001). <https://doi.org/10.1023/A:1010933404324>
4. Chaunzwa, T.L. et al.: Deep learning classification of lung cancer histology using CT images. *Sci. Rep.* **11**(1), 5471 (2021). <https://doi.org/10.1038/s41598-021-84630-x>
5. Clark, K., et al.: The cancer imaging archive (TCIA): Maintaining and operating a public information repository. *J. Digit Imaging.* **26**(6), 1045–1057 (2013). <https://doi.org/10.1007/s10278-013-9622-7>
6. Cortes, C., Vapnik, V.: Support-vector networks. *Mach. Learn.* **20**(3), 273–297 (1995). <https://doi.org/10.1007/BF00994018>
7. Dou, Q., et al.: Multilevel contextual 3-D CNNs for false positive reduction in pulmonary nodule detection. *IEEE Trans. Biomed. Eng.* **64**(7), 1558–1567 (2017). <https://doi.org/10.1109/TBME.2016.2613502>
8. Feng, Y., et al.: GVCNN: Group-view convolutional neural networks for 3D shape recognition. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 264–272 (2018). <https://doi.org/10.1109/CVPR.2018.00035>
9. Guo, Y., et al.: Histological subtypes classification of lung cancers on CT images using 3D deep learning and radiomics. *Acad. Radiol.* **28**(9), e258–e266 (2021). <https://doi.org/10.1016/j.acra.2020.06.010>
10. He, K., et al.: Deep Residual Learning for Image Recognition. <http://arxiv.org/abs/1512.03385> (2015). <https://doi.org/10.48550/arXiv.1512.03385>
11. Li, C., et al.: Multi-view mammographic density classification by dilated and attention-guided residual learning. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **18**(3), 1003–1013 (2021). <https://doi.org/10.1109/TCBB.2020.2970713>

12. Li, S., et al.: Adaptive multimodal fusion with attention guided deep supervision net for grading hepatocellular carcinoma. *IEEE J. Biomed. Health Inform.* **26**(8), 4123–4131 (2022). <https://doi.org/10.1109/JBHI.2022.3161466>
13. Marentakis, P., et al.: Lung cancer histology classification from CT images based on radiomics and deep learning models. *Med. Biol. Eng. Comput.* **59**(1), 215–226 (2021). <https://doi.org/10.1007/s11517-020-02302-w>
14. Meng, Z., et al.: MSMFN: an ultrasound based multi-step modality fusion network for identifying the histologic subtypes of metastatic cervical lymphadenopathy. In: *IEEE Transactions on Medical Imaging*, pp. 1–1 (2022). <https://doi.org/10.1109/TMI.2022.3222541>
15. Pereira, T. et al.: Comprehensive perspective for lung cancer characterisation based on AI solutions using CT images. *J. Clin. Med.* **10**(1), 118 (2021). <https://doi.org/10.3390/jcm10010118>
16. Sahu, P., et al.: A lightweight multi-section CNN for lung nodule classification and malignancy estimation. *IEEE J. Biomed. Health Inform.* **23**(3), 960–968 (2019). <https://doi.org/10.1109/JBHI.2018.2879834>
17. Sedrez, J.A., et al.: Non-invasive postural assessment of the spine in the sagittal plane: a systematic review. *Motricidade* **12**(2), 140–154 (2016). <https://doi.org/10.6063/motricidade.6470>
18. Su, H., et al.: Multi-view convolutional neural networks for 3D shape recognition. In: *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 945–953 (2015). <https://doi.org/10.1109/ICCV.2015.114>
19. Su, R., et al.: Identification of expression signatures for non-small-cell lung carcinoma subtype classification. *Bioinformatics* **36**(2), 339–346 (2019). <https://doi.org/10.1093/bioinformatics/btz557>
20. Tomassini, S., et al.: Lung nodule diagnosis and cancer histology classification from computed tomography data by convolutional neural networks: a survey. *Comput. Biol. Med.* **146**, 105691 (2022). <https://doi.org/10.1016/j.compbiomed.2022.105691>
21. Wang, J., et al.: UASSR: unsupervised arbitrary scale super-resolution reconstruction of single anisotropic 3D images via disentangled representation learning. In: Wang, L. et al. (eds.) *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, pp. 453–462 Springer Nature Switzerland, Cham (2022). https://doi.org/10.1007/978-3-031-16446-0_43
22. Wu, X., et al.: Deep learning-based multi-view fusion model for screening 2019 novel coronavirus pneumonia: a multicentre study. *Eur. J. Radiol.* **128**, 109041 (2020). <https://doi.org/10.1016/j.ejrad.2020.109041>
23. Yanagawa, M., et al.: Diagnostic performance for pulmonary adenocarcinoma on CT: comparison of radiologists with and without three-dimensional convolutional neural network. *Eur. Radiol.* **31**(4), 1978–1986 (2021). <https://doi.org/10.1007/s00330-020-07339-x>
24. Zellinger, W., et al.: Central Moment Discrepancy (CMD) for Domain-Invariant Representation Learning. <http://arxiv.org/abs/1702.08811> (2019)
25. Zhang, N., et al.: Circular RNA circSATB2 promotes progression of non-small cell lung cancer cells. *Mol. Cancer*. **19**(1), 101 (2020). <https://doi.org/10.1186/s12943-020-01221-6>