Pretraining on Chronic Lung Inflammatory Disease Datasets to Enhance Indeterminant Lung Cancer Classification using Masked Autoencoders

Axel H.P. Masquelin[1][0000−0002−9412−0390] and Raúl San José Estépar[1][0000−0002−3677−1996]

Brigham and Women's Hospital, Department of Radiology,
399 Revolution Drive, Somerville, MA, USA 02145
amasquelin@bwh.harvard.edu

Abstract. Lung cancer remains the leading cause of cancer-related mortality in the United States, despite the adoption of low-dose computed tomography (LDCT) and updated screening guidelines from the United States Preventive Service Task Force (USPSTF) [19]. Limited infrastructure and financial costs continue to hinder widespread LDCT adoption, while the increasing detection of indeterminate pulmonary nodules (4–20 mm) challenges accurate diagnosis and clinical decision-making. We address these limitations by pretraining masked autoencoders (MAE) on the COPDGene dataset, which captures chronic lung inflammatory disease features. Emphysema and airway disease, two distinct subtypes of COPD, are pathophysiological manifestations of chronic lung inflammation [4, 15]. Incorporating these features may enhance the model's ability to distinguish between malignant and benign pulmonary nodules. By exploring multiple masking strategies, we optimize network attention on parenchymal and perinodular features, improving the extraction of relevant image biomarkers. Our results demonstrate that pretraining on the COPDGene dataset using random masking (r-masking) achieves superior classification performance, with a sensitivity of 88.79%, specificity of 86.27%, and an AUC of 0.931, when compared to self-pretraining on National Lung Cancer Screening Trial (NLST), and supervised learning on NLST. This highlights the importance of leveraging chronic disease datasets for self-supervised learning and underscores the potential of MAE-based approaches to improve nodule classification in clinical settings. Code available at https://github.com/axemasquelin/RegionalMAE

Keywords: Transfer Learning · Masked Autoencoder · Non-Small Cell Lung Cancer.

1 Introduction

Advancements in artificial intelligence have led to strong optimism about improving the early detection and intervention of cancer. The World Health Organization (WHO) defines early diagnosis as the timely diagnosis of disease before progressing to advanced stages through detecting early pathophysiological development and disease symptoms [16]. In the case of lung cancer, the adoption of low-dose computed tomography (LDCT) as the primary screening modality has led to a 20% reduction in mortality [19].

However, adherence to these screening guidelines remains low across the United States due to screening costs, lack of infrastructure, and personnel to support increased screening [16]. To address this gap in infrastructure and personnel, deep learning methodologies exploring detection and classification have been developed to varying degrees of success [2]. In all cases, training these models requires large data sources to generalize well to novel data. However, as model size and complexity grow, so does its need for data. To address this challenge, pretaining on large datasets, such as ImageNet, and fine-tuning the network on the target domain became the standard methodology of choice. In the case of medical applications, this approach of training on a source domain and transferring knowledge to the target domain does not guarantee that the features learned strongly overlap with those of the target domain [17]. To address this challenge, self-pretraining using masked autoencoders (MAE) has been explored in order to build more robust features [6, 8, 23]. Unlike pretaining, the source and target domains use the same dataset, allowing the model to extract relevant domain knowledge, such as important structures and pathophysiologies associated with disease states. In addition, this approach has improved the model's generalizability across tasks and reduced the reliance on large annotated datasets to learn general high-level features [6]. While self-pretraining using MAE has shown promising results in medical imaging tasks [7, 23], it has its own challenges. One significant limitation of this approach is that it relies on the same dataset for pretraining and downstream tasks, which can lead to model overfitting and be sensitive to dataset size [20].

Considering this, the proposed work explores the application of pretraining using MAE on a chronic obstructive pulmonary disease (COPD) dataset, COPDGene, for downstream indeterminant pulmonary nodule (4mm-20mm) classification on the National Lung Screening Trial (NLST) dataset [1]. Emphysema, a distinct subtypes of COPD, is a pathophysiological manifestation of chronic lung inflammation and is considered pre-malignant conditions by radiologists when reviewing LDCT [4, 15]. Accordingly, increasing model attention to inflammatory image biomarkers is expected to improve the ability of the model to distinguish between malignant and benign nodules accurately. In addition, various masking strategies will be utilized to explore the impact of modulating network attention on parenchymal and perinodular features when compared to standard masking strategies.

1.1 Related Works

Recent advances in self-supervised learning (SSL) have demonstrated performance improvements over both weakly supervised learning and traditional pretraining methods [7, 9, 12, 17]. Among these, masked image modeling (MIM) has emerged as a particularly effective pretraining strategy for both natural and medical images [8, 23, 24]. MIM approaches leverage the reconstruction of masked regions to encourage the learning of robust and context-aware visual representations. In medical imaging, the use of tailored masking strategies, such as non-overlapping or domain-aware masks, has been shown to further enhance the ability of the model to capture fine-grained, clinically relevant features [21]. In parallel, alternative SSL paradigms such as contrastive learning have also gained traction, particularly in the development of pulmonary foundation models [14] and combined MIM approaches [22].

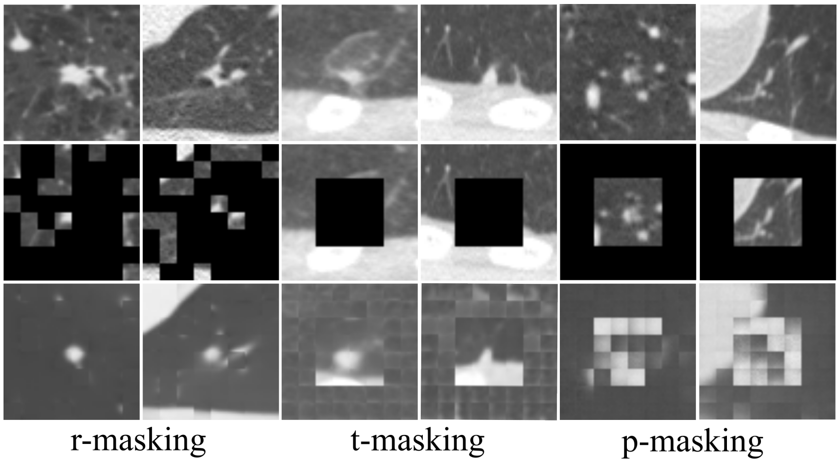

Fig. 1. First row: Window normalized images with COPDGene dataset. Second row: Masked images showing the random masking (r-masking), tumor masking (t-masking), and parenchyma masking (p-masking) strategies. Third row: Reconstructed images from the unmasked patches for the r-masking, t-masking, and p-masking strategies, respectively, where every two columns represent a new masking strategy.

2 Methodology

Figure 1 illustrates the proposed masking strategies, showing the implementation of randomly masking out patches across the image (r-mask), removing tumor and boundary-specific information (t-mask), and the removal of parenchymal information (p-mask) for the pretraining of our Vision Transformer (ViT) encoder.

2.1 Vision Transformer

ViT architectures are the backbone for both the pretraining phase and downstream classification. These architectures comprise a patch embedding layer, positional embedding, and a transformer block for feature extraction. The patch embedding layer transforms a provided image into a set of equal sequences, or patches, based on the provided patch dimension (P). The number of patches (N) is defined as $N = HW/P^2$ for 2D images, where H is the height, and W is the original image's width. These patches, thereafter, become the input for the transformer block of the ViT backbone. A positional embedding layer is utilized to retain positional information following the generation of the patches. Prior work has demonstrated that using a sine-cosine positional embedding improves MAE performance when compared to 1D patch embedding [6]. Lastly, the transformer blocks are created with a depth, D, alternating between multiheaded self-attention layers and multilayer perceptron blocks. To establish a baseline performance for indeterminant nodule classification, pretraining on ImageNet and training from scratch are evaluated. The parameters of our ViT-B-16 model follow prior literature [5], where the embedding dimension is 768, the number of heads is 12, the depth of the encoder is 8, the attention drop rate is 0.1, the drop rate is 0.1 [6].

2.2 Masked Autoencoder

To apply the masking strategies shown in figure 1, the input of the model is divided into equal non-overlapping patches of dimension P. In the case of r-masking, patches are randomly selected as masked. In t-masking, tumor patches are predetermined as the central patches containing all, if not most, of the tumor. In contrast, parenchyma patches are assigned for all other regions. During t-masking, all tumor patches are then assigned as masked while the parenchyma remains visible. Inversely, p-masking sets all parenchyma patches as masked and tasks the MAE to reconstruct them. Visible patches are concatenated with their respective position embedding before the forward function of the ViT to maintain position information related to the visible patches. Before the decoder, the learnable mask tokens are put in the position of the masked patches alongside the full set of tokens, including patch-wise representations, from the encoder. The additional positional embedding to the input tokens of the decoder ensures the restoration of the patches in each given position. Using the reconstructed patches from the MAE decoder, we compare the reconstruction to that of the original patches. The reconstruction loss is only computed across the marked patches.

After pretaining the MAE, a classifier head was appended to the encoder to classify nodules as malignant or benign. The linear classifier takes the encoder embedding and predicts the class output of the nodule. Binary cross-entropy loss is used to train the model. We explore both fine-tuning and linear probing for training the classifier head. In the case of linear probing, only the classifier head is updated, while during fine-tuning the encoder, embedding layer, and classifier are updated, see table 1.

2.3 Datasets and Implementation

COPDGene Phase 1 is a large observational case-control study of COPD that enrolled 10,000 individuals, all of whom underwent CT imaging. Participants had a smoking history of at least 10 pack-years and were classified according to the GOLD criteria into the following COPD stages: 1, 2, 3, or 4 [18, 10]. Additionally, individuals who met the smoking history requirement but had normal spirometry were categorized as ‘Undefined’ (smokers controls). Using a nnUnet to detect the presence of pulmonary nodules, 64x64x64 regions of interest were extracted from each individual. Multiple nodules could occur within one individual, resulting in 51814 nodules being detected. Of which, only 5000 were selected for training the MAE, as the source domain. A 5-fold cross-validation is used to evaluate the performance of the model. A training-validation-test split of 70-10-20 was used. Axial slices (64x64) were randomly selected the region of interest during training. Small rotations between -10 to 10 degrees and random affine were applied to augment the data. Window normalization was applied to all images to enhance the parenchymal tissue signal. A width of 1600 HU and a center at -600 HU was selected.

The National Lung Screening Trial (NLST) dataset was used as the downstream classification task target domain. The low-dose computed tomography branch of the NLST dataset contains 24,517 individuals between the ages of 55 and 74 who had a 30 or more pack-year of cigarette smoking history, were former smokers who had quit in the last 15 years, and were able to lie on their back with their arms raised above their heads [1]. Selecting single solid nodules between 4mm to 20mm in diameter, clearly in the parenchyma and not pleural based, resulted in 3533 individuals, of which 336 were malignant nodules. The size criteria was selected to reduce the influence of diameter on the likelihood of malignancy since solitary nodules with diameters greater than 20mm in diameter are known to be associated with a greater than 50% risk of malignancy [13]. Random downsampling was used to balance the dataset, resulting in 672 images. To augment the data, three central axial slices (64x64) of the nodules were selected if possible, this resulted in a dataset of size 2016. A training-validation-test split of 70-10-20 was used for downstream classification. To allow for proper comparison between MAE pretraining approaches, COPDGene data was set to be equal to NLST.

3 Results

Ablation Studies. Results from the ablation experiment demonstrate that the designed MAE remains stable across various parameters of the study, as shown in Table 1. Parameters such as patch size, number of encoder heads, encoder and decoder depth, loss function, and embedding dimension were evaluated for their role in the model's performance on downstream classification. As seen in table 1a, a patch size of 8 performed slightly better than the standard patch size of 16 typically used for ViT models, with a fine-tuning (ft) accuracy of 90.71% and linear probing (lin) accuracy of 82.24%. Decreasing the number of encoder heads to 8 resulted in the best performance, with 92.39% ft and 83.58% lin accuracy (Table1b). Regarding the loss function, the use of mean squared error with normalization achieved the highest linear probing performance (83.52%),

Table 1. Ablation Experiment on pertaining using COPDGene dataset, reporting fine-tuning (ft) and linear probing (lin) performance.

patch	ft	lin
4	89.50	80.95
8	**90.71**	82.24
16	89.68	**82.30**

(a) Masking Patch Size: size of the masking patch

heads	ft	lin
6	90.30	81.85
8	**92.39**	**83.58**
12	90.71	82.24

(b) Encoder Heads: number of encoder heads

loss	ft	lin
mae (w/o norm)	**92.61**	81.59
mae (w norm)	91.78	82.79
mse (w/o norm)	89.71	82.26
mse (w norm)	91.12	**83.52**

(c) Loss function: loss function performance with and without pixel normalization

depth	ft	lin
4	**92.39**	82.1
6	90.42	81.67
8	90.71	**82.24**
12	92.10	81.15

(d) Encoder Depth: depth of the encoder block following patch

depth	ft	lin
4	92.44	**83.43**
6	**92.84**	81.78
8	90.71	82.24
12	92.56	83.35

(e) Decoder Depth: depth of the decoder block following patch

dim	ft	lin
252	89.22	77.27
516	**90.85**	81.46
768	90.71	82.24
1032	88.97	**83.46**

(f) Embedding: Dimension of the decoder embedding

while mean absolute error without normalization led to the best fine-tuning performance (92.61%) (Figure 1c). For encoder depth, a shallower model with a depth of 4 outperformed deeper configurations, with 92.39% ft and 82.1% lin accuracy (Figure 1d). Similarly, a decoder depth of 12 provided the best overall performance for both fine-tuning and linear probing, seen with 92.56% ft and 83.35% lin accuracy (Table 1e). Finally, increasing the embedding dimension to 1032 led to the highest linear probing accuracy (83.46), while 516 provided a balance between fine-tuning (90.85%) and linear probing (81.46%) performance (Table 1f).

Classification. All experiments reported here followed a 5-fold cross-validation to ensure model performance and stability. As shown in Figure 1, the reconstruction quality of the model varied significantly based on the masking strategy employed and the type of pretraining utilized. Comparing all MAE methodologies, as shown in Table 2, the models pretrained with the COPDGene dataset and using r-masking achieved the highest classification performance across all metrics. This configuration yielded a sensitivity of 88.79% (±2.52), specificity of 86.27% (±3.35), and an AUC of 0.931 (±0.015), highlighting the advantage of regional masking in extracting meaningful representations of premalignant conditions from the COPDGene dataset.

The performance of the ViT-B-16 models without MAE pretraining varied depending on the initialization strategy. The model trained from scratch outperformed the ImageNet pretrained variant, achieving a sensitivity of 71.20% (±8.01), specificity of 78.26% (±4.39), and an AUC of 0.797 (±0.065). In con-

Table 2. Classification and Reconstruction metrics of architectures across masking strategies over 5-fold cross-validation.

Models	**Epochs**		**Sensitivity (%)**	**Specificity (%)**	**AUC**
	MAE	Dx			
ViT-B-16					
scratch	-	100	71.20 ± 8.01	78.26 ± 4.39	0.797 ± 0.065
pretrain	-	100	54.67 ± 10.91	71.81 ± 9.25	0.649 ± 0.032
$MAE_{COPDGene}$					
r-masking	500	100	**88.79 ± 2.52**	**86.27 ± 3.35**	**0.931 ± 0.015**
t-masking	500	100	82.10 ± 5.39	84.55 ± 2.78	0.887 ± 0.026
p-masking	500	100	85.94 ± 3.11	83.09 ± 4.46	0.909 ± 0.017
MAE_{NLST}					
r-masking	500	100	84.24 ± 4.73	85.07 ± 5.46	0.901 ± 0.011
t-masking	500	100	70.14 ± 3.81	73.64 ± 5.16	0.775 ± 0.028
p-masking	500	100	83.33 ± 3.56	79.34 ± 3.56	0.874 ± 0.029

trast, the pretrained ViT-B-16 exhibited lower performance with a sensitivity of 54.67% (±10.91), specificity of 71.81% (±9.25), and an AUC of 0.649 (±0.032), suggesting that domain-specific pretraining is essential for optimal model performance.

When comparing masking strategies, r-masking consistently outperformed both t-masking and p-masking across the two datasets. For pretraining on COPDGene, r-masking showed superior results, with an AUC of 0.931, compared to 0.887 for t-masking and 0.909 for p-masking. A similar trend was observed for pretaining on NLST, where r-masking achieved an AUC of 0.901, compared to 0.775 and 0.874 for t-masking and p-masking, respectively.

The MAE_{NLST} models generally performed slightly worse compared to their $MAE_{COPDGene}$ counterparts, though r-masking still provided the best performance with a sensitivity of 84.24% (±4.73), specificity of 85.07% (±5.46), and an AUC of 0.901 (±0.011).

4 Discussion

Although recent years have led to improved screening guidelines, the NLST dataset still presents clinical limitations due to its relatively healthy population compared to individuals eligible for lung cancer screening in the United States [3]. This creates a gap in generalizability when applying models trained on NLST data to real-world clinical populations with higher disease burdens and comorbidities. Combining non-cancer-specific datasets, like COPDGene, for pretraining offers a crucial advantage by introducing comorbidities common in high-risk populations, such as COPD. This approach allows the model to learn more specific and generalizable chronic inflammatory image biomarkers. These can then be fine-tuned to specific cancer-related tasks on NLST data, as seen with the performance of the $MAE_{COPDGene}$ models when compared to ViT-B-16 in table 2. Additionally, random masking strategies enable the model to avoid learning redundant features by increasing the difficulty of extrapolating nearby patch information [6]. Ensuring that the model learns both perinodular and parenchymal morphologies associated with chronic lung inflammation across the region of interest. When applying a more static masking approach, such as t-masking and p-masking, we can observe a drop in performance for both $MAE_{COPDGene}$ and MAE_{NLST}, as seen in table 2. Furthermore, looking at figure 1, we see that although t-masking strategy improves the quality of reconstruction due to a decrease in the number of pixels it is learning, it fails to improve classification outcomes when compared to p-masking.

Overall, the improved performance of all MAE strategies when compared to standard ViT-B-16 models demonstrates that pretraining on a similar target domain such as COPDGene, or self-pretraining on a medical dataset, allows for more robust and transferable feature representations. Furthermore, the limited success of transfer learning from ImageNet highlights the challenges of applying models pretrained on natural image datasets to medical imaging tasks [17]. Training from scratch or pretraining on a related medical target domain ensures better feature alignment. It leads to more clinically meaningful performance gains, emphasizing the importance of domain-specific pretraining for medical imaging applications.

Nevertheless, the limitations of this study primarily stem from dataset constraints and the nature of baseline comparisons. The baseline models used are not ideal for direct comparison, as modifications to the patch embedding or the size of the region of interest (ROI) fundamentally alter the task. Expanding the ROI to 224x224 likely impacted pretraining performance by increasing the number of anatomical structures the model needed to recognize. Prior work has shown that classification on smaller ROIs often improves performance because the most relevant information for pulmonary nodule classification is concentrated around the nodule itself [11]. This aligns with our findings, where p-masking performed similarly or better than t-masking.

Performance differences between the MAE models trained on COPDGene and NLST datasets can largely be attributed to disparities in dataset size and nodule diversity. The COPDGene dataset included 51,814 potential nodules for pretraining and evaluation, compared to only 3,533 nodules in the NLST dataset. Despite data augmentation by selecting multiple slices from axial view in the NLST dataset, the overall sample size remained insufficient. This data imbalance and lack of nodule diversity likely contributed to the observed performance gap, suggesting that the strict inclusion criteria for the NLST dataset may have limited its representativeness and generalizability.

Moreover, the classification task focused exclusively on solitary pulmonary nodules, tas the incident of cancer ranges from 10 to 70 percent [13]. However, this focus excludes other clinically significant findings that radiologists must evaluate. Expanding the classification task to include ground-glass pulmonary nodules and other alternative findings could lower performance metrics but would significantly enhance the model's clinical applicability and utility.

5 Conclusion

In conclusion, we have demonstrated that MAE pre-training on datasets containing premalignant conditions improves the classification accuracy of the model when compared to both self-pretraining, ImageNet pretraining, and training from scratch. Furthermore, random masking strategies ensure that the model learns robust clinically relevant features when compared to alternative masking strategies. Our approach leverages COPDGene, a dataset with a high prevalence of multimorbidity and compromised lung function, in order to capture the complexity of real-world screening populations while ensuring increased data diversity.

Acknowledgments. This project was funded by the NCI K00CA274713. The content presented herein is solely the responsibility of the authors and does not necessarily represent the official views of the NCI.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. The National Lung Screening Trial: Overview and Study Design. Radiology **258**(1), 243–253 (Jan 2011). https://doi.org/10.1148/radiol.10091808, publisher: Radiological Society of North America
2. Atmakuru, A., Chakraborty, S., Faust, O., Salvi, M., Datta Barua, P., Molinari, F., Acharya, U.R., Homaira, N.: Deep learning in radiology for lung cancer diagnostics: A systematic review of classification, segmentation, and predictive modeling techniques. Expert Systems with Applications **255**, 124665 (Dec 2024). https://doi.org/10.1016/j.eswa.2024.124665
3. Braithwaite, D., Karanth, S., Slatore, C.G., Yang, J.J., Tammemagi, M., Gould, M.K., Silvestri, G.A.: Burden of Comorbid Conditions Among Individuals Screened for Lung Cancer. JAMA Health Forum **6**(2), e245581 (Feb 2025). https://doi.org/10.1001/jamahealthforum.2024.5581
4. Carr, L.L., Jacobson, S., Lynch, D.A., Foreman, M.G., Flenaugh, E.L., Hersh, C.P., Sciurba, F.C., Wilson, D.O., Sieren, J.C., Mulhall, P., Kim, V., Kinsey, C.M., Bowler, R.P.: Features of COPD as Predictors of Lung Cancer. Chest **153**(6), 1326–1335 (Jun 2018). https://doi.org/10.1016/j.chest.2018.01.049
5. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale (Jun 2021). https://doi.org/10.48550/arXiv.2010.11929, arXiv:2010.11929 [cs]
6. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R.: Masked Autoencoders Are Scalable Vision Learners (Dec 2021), arXiv:2111.06377 [cs]

7. Huang, H., Wu, R., Li, Y., Peng, C.: Self-Supervised Transfer Learning Based on Domain Adaptation for Benign-Malignant Lung Nodule Classification on Thoracic CT. IEEE Journal of Biomedical and Health Informatics **26**(8), 3860–3871 (Aug 2022). https://doi.org/10.1109/JBHI.2022.3171851
8. Kraus, O., Kenyon-Dean, K., Saberian, S., Fallah, M., McLean, P., Leung, J., Sharma, V., Khan, A., Balakrishnan, J., Celik, S., Sypetkowski, M., Cheng, C.V., Morse, K., Makes, M., Mabey, B., Earnshaw, B.: Masked Autoencoders are Scalable Learners of Cellular Morphology (Nov 2023), arXiv:2309.16064 [cs]
9. Liu, Z., Li, W., Cui, Y., Chen, X., Pan, X., Ye, G., Wu, G., Liao, Y., Volmer, L., Wee, L., Dekker, A., Han, C., Liu, Z., Shi, Z.: Label-efficient transformer-based framework with self-supervised strategies for heterogeneous lung tumor segmentation. Expert Systems with Applications **269**, 126364 (Apr 2025). https://doi.org/10.1016/j.eswa.2024.126364
10. Maselli, D.J., Bhatt, S.P., Anzueto, A., Bowler, R.P., DeMeo, D.L., Diaz, A.A., Dransfield, M.T., Fawzy, A., Foreman, M.G., Hanania, N.A., Hersh, C.P., Kim, V., Kinney, G.L., Putcha, N., Wan, E.S., Wells, J.M., Westney, G.E., Young, K.A., Silverman, E.K., Han, M.K., Make, B.J.: Clinical Epidemiology of COPD. Chest **156**(2), 228–238 (Aug 2019). https://doi.org/10.1016/j.chest.2019.04.135
11. Masquelin, A.H., Alshaabi, T., Cheney, N., Estépar, R.S.J., Bates, J.H.T., Kinsey, C.M.: Perinodular Parenchymal Features Improve Indeterminate Lung Nodule Classification. Academic Radiology **30**(6), 1073–1080 (Jun 2023). https://doi.org/10.1016/j.acra.2022.07.001
12. Matsoukas, C., Haslum, J.F., Sorkhei, M., Söderberg, M., Smith, K.: What Makes Transfer Learning Work for Medical Images: Feature Reuse & Other Factors. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 9215–9224 (Jun 2022). https://doi.org/10.1109/CVPR52688.2022.00901, iSSN: 2575-7075
13. Ost, D.E., Gould, M.K.: Decision Making in Patients with Pulmonary Nodules. American Journal of Respiratory and Critical Care Medicine **185**(4), 363–372 (Feb 2012). https://doi.org/10.1164/rccm.201104-0679CI, publisher: American Thoracic Society - AJRCCM
14. Pai, S., Bontempi, D., Hadzic, I., Prudente, V., Sokač, M., Chaunzwa, T.L., Bernatz, S., Hosny, A., Mak, R.H., Birkbak, N.J., Aerts, H.J.W.L.: Foundation model for cancer imaging biomarkers. Nature Machine Intelligence **6**(3), 354–367 (Mar 2024). https://doi.org/10.1038/s42256-024-00807-9, publisher: Nature Publishing Group
15. Parris, B.A., O'Farrell, H.E., Fong, K.M., Yang, I.A.: Chronic obstructive pulmonary disease (COPD) and lung cancer: common pathways for pathogenesis. Journal of Thoracic Disease **11**(Suppl 17), S2155–S2172 (Oct 2019). https://doi.org/10.21037/jtd.2019.10.54
16. Poon, C., Wilsdon, T., Sarwar, I., Roediger, A., Yuan, M.: Why is the screening rate in lung cancer still low? A seven-country analysis of the factors affecting adoption. Frontiers in Public Health **11**, 1264342 (2023). https://doi.org/10.3389/fpubh.2023.1264342
17. Raghu, M., Zhang, C., Kleinberg, J., Bengio, S.: Transfusion: Understanding Transfer Learning for Medical Imaging (Oct 2019). https://doi.org/10.48550/arXiv.1902.07208, arXiv:1902.07208 [cs]
18. Regan, E.A., Hokanson, J.E., Murphy, J.R., Make, B., Lynch, D.A., Beaty, T.H., Curran-Everett, D., Silverman, E.K., Crapo, J.D.: Genetic Epidemiology of COPD (COPDGene) Study Design. COPD **7**(1), 32–43 (Feb 2010). https://doi.org/10.3109/15412550903499522

19. US Preventive Services Task Force, Krist, A.H., Davidson, K.W., Mangione, C.M., Barry, M.J., Cabana, M., Caughey, A.B., Davis, E.M., Donahue, K.E., Doubeni, C.A., Kubik, M., Landefeld, C.S., Li, L., Ogedegbe, G., Owens, D.K., Pbert, L., Silverstein, M., Stevermer, J., Tseng, C.W., Wong, J.B.: Screening for Lung Cancer: US Preventive Services Task Force Recommendation Statement. JAMA **325**(10), 962–970 (Mar 2021). https://doi.org/10.1001/jama.2021.1117
20. Wolf, D., Payer, T., Lisson, C.S., Lisson, C.G., Beer, M., Götz, M., Ropinski, T.: Self-supervised pre-training with contrastive and masked autoencoder methods for dealing with small datasets in deep learning for medical imaging. Scientific Reports **13**(1), 20260 (Nov 2023). https://doi.org/10.1038/s41598-023-46433-0, publisher: Nature Publishing Group
21. Xie, Y., Gu, L., Harada, T., Zhang, J., Xia, Y., Wu, Q.: Rethinking masked image modelling for medical image representation. Medical Image Analysis **98**, 103304 (Dec 2024). https://doi.org/10.1016/j.media.2024.103304
22. Zhao, T., Yue, Y., Sun, H., Li, J., Wen, Y., Yao, Y., Qian, W., Guan, Y., Qi, S.: MAEMC-NET: a hybrid self-supervised learning method for predicting the malignancy of solitary pulmonary nodules from CT images. Frontiers in Medicine **12** (Feb 2025). https://doi.org/10.3389/fmed.2025.1507258, publisher: Frontiers
23. Zhou, L., Liu, H., Bae, J., He, J., Samaras, D., Prasanna, P.: Self Pre-Training with Masked Autoencoders for Medical Image Classification and Segmentation. In: 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI). pp. 1–6 (Apr 2023). https://doi.org/10.1109/ISBI53787.2023.10230477, iSSN: 1945-8452
24. Zhuang, J., Wu, L., Wang, Q., Fei, P., Vardhanabhuti, V., Luo, L., Chen, H.: MiM: Mask in Mask Self-Supervised Pre-Training for 3D Medical Image Analysis (Jan 2025). https://doi.org/10.48550/arXiv.2404.15580, arXiv:2404.15580 [cs] version: 2