An Interpretable AI Model to Support CT-Detected Lung Nodule Classification

Liam Connor[1], Michael Scutari[1], Luke Moffett[1], Jon Donnelly[1], Zhicheng Guo[1], Fides Schwartz[2], Joseph Lo[1], and Cynthia Rudin[1]

[1] Duke University, Durham NC 27705, USA
[2] Brigham and Women's Hospital, Harvard University, Boston MA 02115, USA

Abstract. Deep learning models have shown remarkable performance in predicting the malignancy of lung nodules via CT scans, offering a promising step towards automated lung cancer diagnosis. However, the best automated methods for categorizing lung nodule malignancy are black-boxes, making them difficult to trust in clinical practice. To overcome this challenge, we introduce a new interpretable lung nodule malignancy classification model, Prototypical Pulmonary Net (ProtoPulmNet), that is comparable to black-box AI models. ProtoPulmNet is a cascaded model with two stages. The first stage is a simple generalized additive model (GAM) trained on tabular data. The GAM uses reasoning that is consistent with easily understood clinical rules (e.g., classifying small nodules as benign). When clinical rules are insufficient, the second stage in the cascade is invoked. The patient's image data is provided to an interpretable prototypical-part network (PPN). This network is trained to learn cases that are challenging to classify using tabular features alone. To make predictions, the PPN makes comparisons between the patient's images and prototypical training examples. ProtoPulmNet's cascaded architecture enables far fewer image comparisons than other interpretable deep learning methods but maintains competitive performance.

Keywords: Interpretability · Lung Nodule Classification · Deep Learning.

1 Introduction

Large, randomized clinical trials like the National Lung Screening Trial (NLST) have demonstrated that screening individuals at high risk of lung cancer can reduce lung cancer mortality [22, 8]. Radiologists currently evaluate chest computed tomography (CT) exams for lung cancer using the Lung-RADS system [9], a lexicon based on clinically validated heuristics that requires radiologists' judgment to extract features. The Lung-RADS system guides clinical workup decisions in a domain where both false negatives and false positives carry risk; false negatives may result in delays in treatment, and false positives may result in unnecessary imaging or biopsy, which is associated with non-trivial risk [7]. Accurate, AI-assisted nodule diagnosis directly from chest CT may provide a pathway to better navigate this tradeoff.

Recently, deep-learning approaches have demonstrated strong performance in detecting and classifying the malignancy of lung nodules [2, 11, 24, 25]. This performance has motivated public grand challenges, such as LUNA25 [18], which has created further computer vision nodule malignancy classifiers. These classifiers, however, sacrifice the simple rule-based nature of Lung-RADS and ultimately produce black-box models that are difficult to incorporate into clinical practice. Interpretable deep-learning models present a promising alternative to produce strong performance while fitting more naturally into clinical practice.

In this work, we introduce an interpretable AI model, ProtoPulmNet, that is designed to aid (rather than replace) clinical radiologists in diagnosis of screening CT-identified lung nodules. ProtoPulmNet is a two-stage cascade model that uses a combination of clinical rules and prototypical-case comparisons. First, a generalized additive model (GAM) classifies the nodule by its diameter. When this information is insufficient, ProtoPulmNet cascades to a prototypical-part

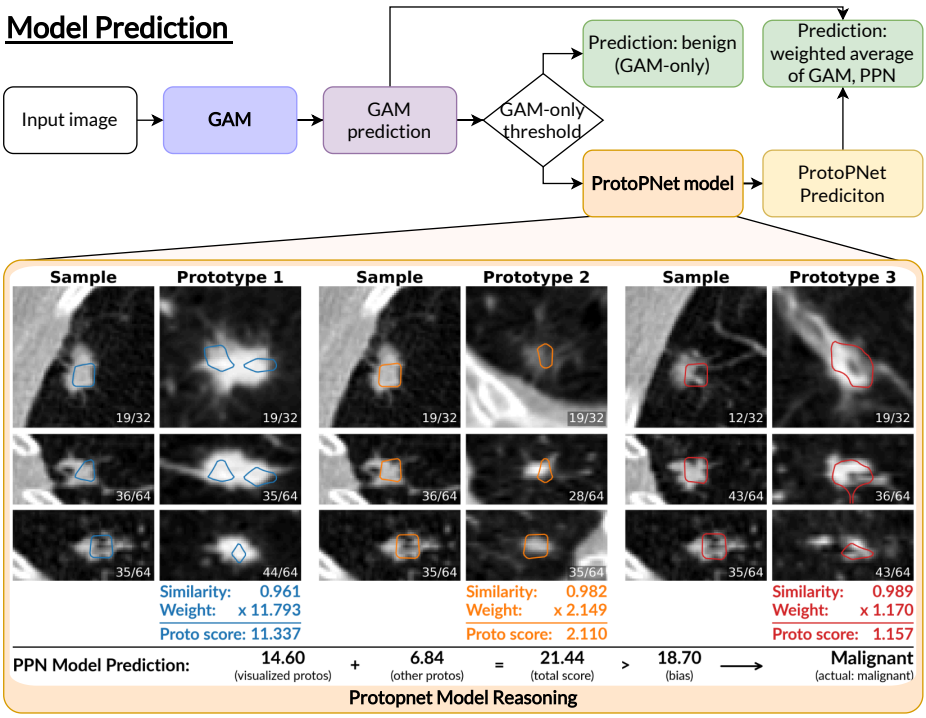

Fig. 1. Model prediction process. *Flowchart (top):* Model cascade flowchart, showing how a sample is first evaluated by the GAM and then evaluated by the PPN if necessary. *ProtoPNet Model Reasoning (bottom):* The PPN model reasoning process for a malignant nodule illustrated by the three prototypes with highest activation for the sample. Contour lines representing the 90^{th} percentile of activation. Corresponding colored lines represent parts of the sample and prototype that the model thinks resemble each other. Rows depict the axial, coronal, and sagittal slice with highest activation.

network (PPN). The PPN makes predictions by visually comparing the nodule to prototypical malignant nodules (shown in Figure 1) learned from training data. This approach is inspired by ProtoPNet [3] and its successor models [5, 10, 14, 15, 19, 20, 26, 27], including applications to medical imaging tasks [1, 29, 30, 28, 21]. In a break from prior PPNs, all of ProtoPulmNet's prototypes contribute only positive evidence to classification. Because it is classifying malignancy, all prototypes are malignant nodules. Moreover, to encourage the PPN to learn nodules that are misclassified by size alone, we reweight samples on errors of the GAM. This architecture allows ProtoPulmNet to predict the malignancy of many nodules in LUNA25 using only a simple GAM. For the remaining samples that use deep learning for prediction, ProtoPulmNet's predictions are based on interpretable comparisons to learned prototypical malignant nodules.

Contributions of this study: (1) Demonstrate the first interpretable, prototype-based AI model for lung nodule malignancy classification with performance similar to black-box AI models. (2) Design a new architecture and training regime that leverages deep learning image comparisons for reasoning while minimizing the architecture's dependence on deep learning. (3) Develop a novel binary classification method for PPNs using only positive evidence, preventing reasoning from the absence of counterexamples.

2 Methods

Let $\mathcal{D} = \{(X_i, z_i, y_i)\}_{i=1}^{N}$ denote the dataset, where $X_i \in \mathbb{R}^{D\times H\times W}$ is a 3D CT region-of-interest (ROI) centered on a candidate lung nodule, $z_i \in \mathbb{R}^s$ are structured clinical features of the nodule, and $y_i \in \{0, 1\}$ is the malignancy label.

ProtoPulmNet consists of a two-stage cascaded architecture that combines a generalized additive model (GAM) on tabular features with a prototypical-part image model. Figure 1 provides an overview of the model structure. When passed a sample (X_i, z_i, y_i), the model first classifies the sample with the GAM. The GAM, $g : \mathbb{R}^s \to \mathbb{R}$, operates on the sample's structured clinical features and produces a malignancy probability, $q_\phi(z_i)$. If the resulting malignancy probability is below a selected threshold, $\beta \in \mathbb{R}$, then the sample is classified by the GAM alone. Otherwise, if $q_\phi(z_i) > \beta$, then the sample's CT data X_i is passed to the PPN. The PPN compares the image to known prototypes and returns a malignancy probability $q_\psi(X_i)$. The final model prediction is a weighted average of the GAM and PPN model probabilities.

$$p_{\text{joint}}(X, z) = \begin{cases} \alpha\, q_\psi(X) + (1-\alpha)\, q_\phi(z) & \text{if } q_\phi(z) \geq \beta \\ q_\phi(z) & \text{else} \end{cases}. \tag{1}$$

The parameter α is calculated over the validation set to maximize performance.

2.1 GAM Architecture and Training

The GAM captures clinically interpretable decision rules on tabular nodule data. The GAM, g, produces logits from the tabular data, ℓ_{gam}, and a malignancy

probability $q_\phi(z_i)$. Because this model is a GAM, the relationship between each input feature and the outcome can be visualized as a 1D shape function. The GAM is trained to minimize cross-entropy and remains frozen during PPN training. We use binned nodule diameter as the only tabular clinical feature in ProtoPulmNet.

2.2 Prototypical-Part Model Architecture and Training

PPN Model Architecture. The 3D PPN architecture performs case-based reasoning on CT image data. A 3D convolutional backbone $f_\theta : \mathbb{R}^{D\times H\times W} \to \mathbb{R}^{C\times D'\times H'\times W'}$ maps the input volume to a latent feature representation. Here, D', H' and W' denote the spatial dimensions of the latent representation. We calculate similarities between latent features of the sample $f_\theta(X)$ and each member of a set of learned prototypes, which we denote by $P = \{p_m\}_{m=1}^{M}$ with $p_m \in \mathbb{R}^C$, using cosine similarity. This produces activation maps $A \in \mathbb{R}^{M\times D'\times H'\times W'}$, where $a_{m,d,h,w}$ is large if prototype m resembles the input at location (d, h, w). Let s_m denote the maximum activation of prototype m on X across all spatial locations. To classify a sample, s_m values are aggregated by a linear prediction head, $h : \mathbb{R}^M \to \mathbb{R}$, which produces an image-based logit ℓ_{ppn} and a malignancy probability $q_\psi(X) = \sigma(\ell_{\text{ppn}})$. In contrast to prior work, we learn only positive class (malignant) prototypes. We therefore introduce a bias term, h_b, on the linear prediction head to enable classification with only positive evidence.

Bias Initialization. If we naively initialized h_b to 0, all predictions of the PPN would land in a low-gradient region of cross entropy loss because we have only positive class prototypes. To mitigate this problem, we first randomly initialize prototype candidates, extract activations from the training set, and then optimize the bias term over cross-entropy loss with all other values held constant.

PPN Objective. The PPN is trained using a weighted objective that combines classification and prototype-regularization terms. The total loss, $\mathcal{L}$, is given in Eq. 2. A detailed explanation of loss terms is given below. All coefficients $\lambda_\cdot$ correspond to the hyperparameters used in training.

$$\mathcal{L} = \lambda_{\text{wbce}}\mathcal{L}_{\text{wbce}} + \lambda_{\text{clst}}\mathcal{L}_{\text{clst}} + \lambda_{\text{sep}}\mathcal{L}_{\text{sep}} + \lambda_{\text{fa}}\mathcal{L}_{\text{fa}} + \lambda_{\text{ortho}}\mathcal{L}_{\text{ortho}} + \lambda_{\ell_1}\mathcal{L}_{\ell_1}. \quad (2)$$

To learn from samples with high GAM classification error, we apply binary cross-entropy to the malignancy logit $\ell_{\text{ppn},i}$ and weight samples by w_i such that

$$\mathcal{L}_{wbce} := \frac{1}{N}\sum_{i=1}^{N} w_i \left[y_i \log(q_\psi(X_i)) + (1 - y_i)\log(1 - q_\psi(X_i))\right],$$

where $w_i := \exp(-\ell_{\text{gam},i}(2y_i - 1))$.

To encourage prototype diversity, we implement orthogonality loss from [27], notated as $\mathcal{L}_{ortho}$. To encourage sparse connections, we add an ℓ_1 penalty on the weights of the prototype-to-class linear layer, $\mathcal{L}_{\ell_1}$ as in [3].

We adapt ProtoPNet-style [3] clustering and separation losses, which encourage each sample to activate prototypes of the correct class and suppress activation of incorrect-class prototypes, to the positive-reasoning-only setting. Let

$\mathcal{I}^+ = \{i \in \mathcal{D} : y_i = 1\}$ be the set of positive samples, and $\mathcal{I}^- = \{i \in \mathcal{D} : y_i = 0\}$ be the set of negative samples. The cluster and separation losses are defined as:

$$\mathcal{L}_{\text{clst}} = -\frac{1}{\sum_{i \in \mathcal{I}^+} w_i} \sum_{i \in \mathcal{I}^+} w_i \max_m s_{i,m}, \qquad \mathcal{L}_{\text{sep}} = \frac{1}{|\mathcal{I}^-|} \sum_{i \in \mathcal{I}^-} w_i \max_m s_{i,m}.$$

Finally, to incentivize prototype localization, we penalize prototype activations outside the nodule segmentation by adapting fine-annotation loss from breast imaging PPNs [1, 12, 29]. Let $\Omega_i \in [0,1]^{D' \times H' \times W'}$ denote a mask indicating valid voxels on which to activate for sample i downsampled to the activation map resolution and let $A_i \in \mathbb{R}^{M \times D' \times H' \times W'}$ be the tensor of prototype activations for sample i. The fine-annotation loss penalizes off-mask activations via

$$\mathcal{L}_{\text{fa}} = \frac{1}{NM} \sum_{i=1}^{N} \sum_{m=1}^{M} \left\| (1 - \Omega_i) \odot A_{i,m} \right\|_2 .$$

3 Experiments

3.1 Experimental Setup

Dataset. We train and evaluate ProtoPulmNet using LUNA25 [16, 17]. LUNA25 contains low-dose CT scans of 6,163 nodules (555 malignant, 5,608 benign) from 2,120 patients who participated in the National Lung Cancer Screening Trial (NLST) between 2002 and 2004. A small subset of samples (<5%) were excluded due to preprocessing inconsistencies, resulting in a final experimental cohort of 5,883 samples. We split the dataset into training, validation, and test sets using an 80/10/10 ratio. Prior to model input, ROIs were resampled to 1.0mm × 0.5mm × 0.5mm for $D \times H \times W$ using linear interpolation, cropped to a fixed size of $32 \times 64 \times 64$ voxels, and windowed to HU values from -1000 to 400. Training data were resampled for class balance, and image data was augmented with random cropping, flips, rotations, and noise during training.

We use nodule diameter as the sole clinical tabular feature. Nodule volume is calculated using segmentations obtained from an existing KNN segmentation pipeline [23]. We estimated nodule diameter from volume, assuming spherical nodules. This procedure was chosen to benefit from the increased discriminatory power of volume for malignancy classification [6] while aligning with Lung-RADS clinical guidelines. The diameter is encoded using cumulative binary encoding (e.g., diameter $\leq$ 2mm, diameter $\leq$ 4mm, etc).

GAM Implementation. We initialized the model with positive weights and a negative bias to encourage monotonicity in nodule size. We used binary cross-entropy as a loss function. The GAM was trained using SGD until the validation set accuracy failed to improve for five consecutive epochs. The resulting GAM achieved a test AUC of 0.66 (95% CI: [0.57, 0.73]) using this approach.

PPN Model Implementation. We initialized the 3D convolutional backbone f_θ with a pretrained MedicalNet ResNet-50 [4] model. To preserve spatial information for prototype localization, the stride of the initial convolution layer was

reduced from two to one, and max pooling was removed, resulting in a latent spatial dimension of $D' \times H' \times W' = 4 \times 8 \times 8$. After bias initialization, we trained the PPN model according to the regime described by [3]. We used Adam to minimize the overall loss given by Equation (2). The model was trained until the joint model AUC on the validation set did not improve for three consecutive project phases [13]. Hyperparameters were selected via Bayesian optimization to maximize AUC on the validation split across 34 runs. The final model was trained for 17.5 hours on one NVIDIA A6000 GPU.

Cascade Model Threshold and Weighting. We describe GAM certainty thresholds, β, in terms of nodule diameter. This is possible because our GAM predictions are monotonic in and solely dependent on nodule volume. We evaluate ProtoPulmNet at two diameter thresholds for GAM-only classification, $z_{\text{dia}} \in \{6.0\text{mm}, 8.0\text{mm}\}$, which were informed by the Lung-RADS 4A and 4B criteria. In a clinical setting, the choice of β would depend on empirical study of the trade-off between time saved by radiologists interpreting the simple GAM predictions and the more powerful PPN predictions. For each β, we determined an α by optimizing for AUC on the validation set.

Comparison Models. We consider two baseline models for comparison using the ResNet50 MedicalNet backbone: (1) a black-box model with a linear prediction head, and (2) a deep-kNN model, which performs comparisons in the latent space. The number of neighbors k in the kNN was selected to minimize k subject to classification AUC greater-than-or-equal to ProtoPulmNet's. To find optimal black-box model training parameters, we ran a Bayesian hyperparameter optimization for approximately 120 clock hours on an NVIDIA A6000 GPU.

3.2 Experimental Results

	Performance	**Interpretability**		
Method	AUC (95% CI)	Min Comp *1 Sample*	Max Comp *1 Sample*	Total Comp *All Samples*
Black-box	**0.86 (0.80, 0.91)**	-	-	-
kNN	0.83 (0.76, 0.89)	12	12	4,190
ProtoPulmNet-6mm	0.83 (0.77, 0.88)	**0**	**9**	**9**
ProtoPulmNet-8mm	0.80 (0.73, 0.86)	**0**	**9**	**9**

Table 1. *Performance*: Classification AUC (95% CI) of ProtoPulmNet compared to black-box and kNN. We evaluate our model at two thresholds, 6.0mm, 8.0mm, to illustrate threshold tradeoffs. *Interpretability*: The number of comparisons used by ProtoPulmNet and its baseline kNN. *1 Sample* columns quantify how many prototypes/neighbors are necessary for inference on one sample, while the *Total Comp* column quantifies how many prototypes/neighbors would need to be inspected to understand the entire model. ProtoPulmNet's efficient cascade requires 0 comparisons for some samples.

ProtoPulmNet performs similarly to black-box baselines. Table 1 (left) compares the predictive performance of ProtoPulmNet to black-box and kNN baselines. ProtoPulmNet-6mm achieved an AUC of 0.83 (0.77,0.88), compared to the 0.86 (0.80, 0.91) of the black-box baseline, with a large amount of overlap in the CIs. Notably, ProtoPulmNet adds the benefit of interpretability relative to the black-box model with comparable performance. This result indicates that our interpretable approach can be used in place of black-box models in diagnostic pipelines, improving clinical trust and aiding adoption.

ProtoPulmNet's reasoning is sparser than alternative interpretable approaches. We evaluate model interpretability by quantifying how many visual comparisons each model entails, as shown in Table 1 (bottom). ProtoPulmNet requires nine comparisons to evaluate model reasoning on a given sample, versus the 12 required by an equivalently-performing kNN. This reduces the time required to interpret and verify ProtoPulmNet's reasoning process. Moreover, the same nine examples are used for every sample by ProtoPulmNet, in comparison to the 4,190 potential neighbors under kNN. To inspect the entire model's reasoning, a user must consider nine cases for ProtoPulmNet, compared to 4,190 cases for a kNN, making ProtoPulmNet much simpler to understand.

Up to 27% of samples are classified by the GAM alone with minimal impact on performance. Our cascade architecture permits many samples to be classified using tabular data directly, avoiding the need for visual comparison, as shown in Fig. 2. For the LUNA25 dataset, we are able to classify 12% and 27% of samples with the GAM alone, for the 6.0mm and 8.0mm cutoffs respectively. The reasoning process for classifying these nodules as benign is easily verifiable and interpretable. Additionally, our model architecture enables different thresholds β to be chosen to balance clinician time and model accuracy.

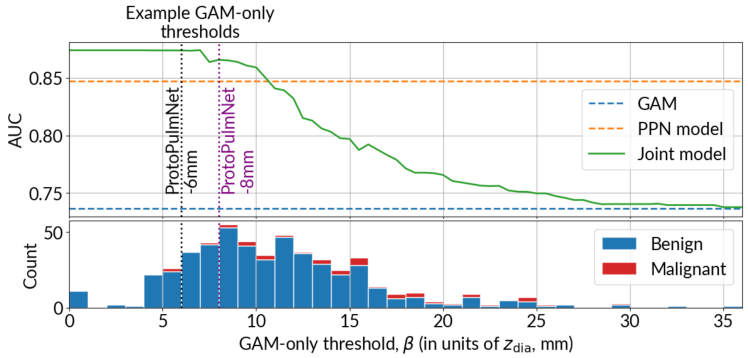

Fig. 2. *Top*: The x-axis represents a diameter threshold z_{dia}: nodules below this threshold are classified by the GAM alone; those above are classified jointly by the GAM and PPN. Moving right, more nodules are handled by the GAM alone, and joint model performance degrades towards the GAM. Vertical lines mark the thresholds evaluated in Table 1. *Bottom*: Diameter and class distribution of the validation set.

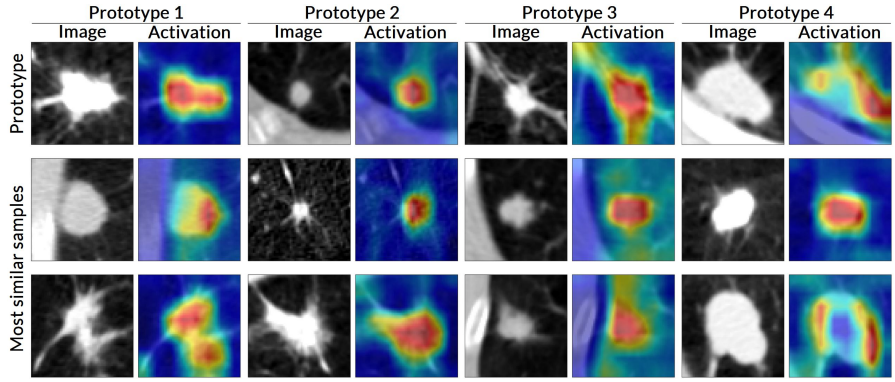

Fig. 3. Malignant prototypes learned by ProtoPulmNet and their self-activation heatmap. The PPN model component classifies samples by judging a sample's similarity with learned prototypes. Activation is calculated as the cosine similarity between the prototype patch and the corresponding image, upsampled from latent space. The top row shows prototypes. The following rows show the two samples with the highest activation for that prototype. The axial slice with highest activation is displayed for each sample. Animated slice views are available in the Supplemental Material.

ProtoPulmNet's prototypes are semantically meaningful. Figure 3 presents four prototypes from ProtoPulmNet alongside their most similar samples. For example, prototype 4 is most similar to large, solid nodules with well-defined margins. Moreover, the explanations from ProtoPulmNet are faithful and simple to understand, as shown in Figure 1. Users simply need to review which prototypes were considered similar to an input to faithfully understand the model reasoning process. In contrast to black-box models, this may enable users to identify and reject cases of confounded reasoning in the clinic.

4 Conclusion

We introduced ProtoPulmNet, an interpretable deep learning model designed to aid radiologists in CT lung nodule malignancy classification. By using a GAM for small nodules and prototypical comparisons for larger nodules, ProtoPulmNet produces interpretable predictions with minimal interpretation overhead. In our study, we found that ProtoPulmNet has similar predictive performance compared to a black-box alternative. In comparison to kNN, ProtoPulmNet requires fewer comparisons and its entire set of nine prototypes can be inspected.

In addition to the limitations inherent in any retrospective study, our study is limited to only one dataset, LUNA25. Additionally, nodule diameters were calculated from segmentation under the assumption of spherality, which may differ systematically from radiologist estimates of diameters. Future work should directly assess the efficacy of radiologists working alongside ProtoPulmNet, and evaluate ProtoPulmNet on external test datasets.

References

1. Barnett, A.J., Schwartz, F.R., Tao, C., Chen, C., Ren, Y., Lo, J.Y., Rudin, C.: A case-based interpretable deep learning model for classification of mass lesions in digital mammography. Nature Machine Intelligence **3**(12), 1061–1070 (2021)
2. Cardoso, M.J., Li, W., Brown, R., Ma, N., Kerfoot, E., Wang, Y., Murrey, B., Myronenko, A., Zhao, C., Yang, D., et al.: Monai: An open-source framework for deep learning in healthcare. arXiv preprint arXiv:2211.02701 (2022)
3. Chen, C., Li, O., Tao, D., Barnett, A., Rudin, C., Su, J.K.: This looks like that: deep learning for interpretable image recognition. Advances in neural information processing systems **32** (2019)
4. Chen, S., Ma, K., Zheng, Y.: Med3d: Transfer learning for 3d medical image analysis. arXiv preprint arXiv:1904.00625 (2019)
5. Donnelly, J., Barnett, A.J., Chen, C.: Deformable Protopnet: An Interpretable Image Classifier Using Deformable Prototypes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10265–10275 (2022)
6. Han, D., Heuvelmans, M.A., Oudkerk, M.: Volume versus diameter assessment of small pulmonary nodules in CT lung cancer screening **6**(1), 52–61. https://doi.org/10.21037/tlcr.2017.01.05, https://pmc.ncbi.nlm.nih.gov/articles/PMC5344834/
7. Heerink, W.J., de Bock, G.H., de Jonge, G.J., Groen, H.J., Vliegenthart, R., Oudkerk, M.: Complication rates of ct-guided transthoracic lung biopsy: meta-analysis. European radiology **27**(1), 138–148 (2017)
8. de Koning, H.J., van Der Aalst, C.M., de Jong, P.A., Scholten, E.T., Nackaerts, K., Heuvelmans, M.A., Lammers, J.W.J., Weenink, C., Yousaf-Khan, U., Horeweg, N., et al.: Reduced lung-cancer mortality with volume ct screening in a randomized trial. New England journal of medicine **382**(6), 503–513 (2020)
9. of Radiology Committee on Lung-RADS, A.C.: Lung-RADS Assessment Categories 2022 (2023), https://www.jacr.org/article/S1546-1440(23)00761-5/fulltext
10. Ma, C., Zhao, B., Chen, C., Rudin, C.: This Looks Like Those: Illuminating Prototypical Concepts Using Multiple Visualizations. Advances in Neural Information Processing Systems **36** (2024)
11. Massion, P.P., Antic, S., Ather, S., Arteta, C., Brabec, J., Chen, H., Declerck, J., Dufek, D., Hickes, W., Kadir, T., et al.: Assessing the accuracy of a deep learning method to risk stratify indeterminate pulmonary nodules. American journal of respiratory and critical care medicine **202**(2), 241–249 (2020)
12. Moffett, L., Barnett, A.J., Donnelly, J., Schwartz, F.R., Trivedi, H., Lo, J., Rudin, C.: Multi-site validation of an interpretable model to analyze breast masses. PLoS One **20**(6), e0320091 (2025)
13. Moffett, L., Willard, F., Machado, M., Mokel, E., Donnelly, J., Guo, Z., Costarino, A., Yang, J., Kim, G., Barnett, A.J., Rudin, C.: Cosine Similarity is Almost All You Need (for Prototypical-Part Models). In: 2026 IEEE winter conference on applications of computer vision (WACV). IEEE (2026)
14. Nauta, M., Jutte, A., Provoost, J., Seifert, C.: This Looks Like That, Because... Explaining Prototypes for Interpretable Image Recognition. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. pp. 441–456. Springer (2021)
15. Nauta, M., Van Bree, R., Seifert, C.: Neural Prototype Trees for Interpretable Fine-Grained Image Recognition. In: Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition. pp. 14933–14943 (2021)

16. Peeters, D., Obreja, B., Antonissen, N., Jacobs, C.: The luna25 challenge: Public training and development set - annotation data (2025). https://doi.org/10.5281/zenodo.14673658, https://doi.org/10.5281/zenodo.14673658
17. Peeters, D., Obreja, B., Antonissen, N., Jacobs, C.: The luna25 challenge: Public training and development set - imaging data (2025). https://doi.org/10.5281/zenodo.14223624, https://doi.org/10.5281/zenodo.14223624
18. Peeters, D., Obreja, B., Antonissen, N., Jacobs, C.: The luna25 challenge: Public training and development set-imaging data (2025)
19. Rymarczyk, D., Struski, Ł., Górszczak, M., Lewandowska, K., Tabor, J., Zieliński, B.: Interpretable Image Classification With Differentiable Prototypes Assignment. In: European Conference on Computer Vision. pp. 351–368. Springer (2022)
20. Rymarczyk, D., Struski, Ł., Tabor, J., Zieliński, B.: ProtoPShare: Prototypical Parts Sharing for Similarity Discovery in Interpretable Image Classification. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. pp. 1420–1430 (2021)
21. Sarangi, P., Agarwal, R., Basu, T.: Protoradnet: Prototypical patches of convolutional features for radiology image classification network. Artificial Intelligence in Medicine p. 103324 (2025)
22. Team, N.L.S.T.R.: Reduced lung-cancer mortality with low-dose computed tomographic screening. New England Journal of Medicine **365**(5), 395–409 (2011)
23. Tushar, F.I.: Peritumoral Expansion Radiomics for Improved Lung Cancer Classification
24. Tushar, F.I., Wang, A., Dahal, L., Samei, E., Harowicz, M.R., Kalpathy-Cramer, J., Lafata, K.J., Tailor, T.D., Rudin, C., Lo, J.Y.: Ai in lung health: Benchmarking detection and diagnostic models across multiple ct scan datasets. arXiv preprint arXiv:2405.04605 (2024)
25. Venkadesh, K.V., Aleef, T.A., Scholten, E.T., Saghir, Z., Silva, M., Sverzellati, N., Pastorino, U., van Ginneken, B., Prokop, M., Jacobs, C.: Prior ct improves deep learning for malignancy risk estimation of screening-detected pulmonary nodules. Radiology **308**(2), e223308 (2023)
26. Wang, C., Liu, Y., Chen, Y., Liu, F., Tian, Y., McCarthy, D., Frazer, H., Carneiro, G.: Learning Support and Trivial Prototypes for Interpretable Image Classification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2062–2072 (2023)
27. Wang, J., Liu, H., Wang, X., Jing, L.: Interpretable Image Recognition by Constructing Transparent Embedding Space. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 895–904 (2021)
28. Wei, Y., Tam, R., Tang, X.: Mprotonet: A case-based interpretable model for brain tumor classification with 3d multi-parametric magnetic resonance imaging. In: Medical Imaging with Deep Learning. pp. 1798–1812. PMLR (2024)
29. Yang, J., Barnett, A.J., Donnelly, J., Kishore, S., Fang, J., Schwartz, F.R., Chen, C., Lo, J.Y., Rudin, C.: Fpn-iaia-bl: a multi-scale interpretable deep learning model for classification of mass margins in digital mammography. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5003–5009 (2024)
30. Yazdani, E., Neizehbaz, A., Karamzade-Ziarati, N., Kheradpisheh, S.R.: Explainable artificial intelligence for pneumonia classification: clinical insights into deformable prototypical part network in pediatric chest x-ray images. Journal of medical imaging and radiation sciences **56**(5), 102023 (2025)