AC-MIL: Weakly-Supervised Atrial LGE-MRI Quality Assessment via Adversarial Concept Disentanglement

K M Arefeen Sultan[1,2], Kaysen Hansen [5], Benjamin Orkild[1,3,4], Alan Morris[1], Eugene Kholmovski[5,6], Erik Bieging[7], Eugene Kwan[3,4], Ravi Ranjan[3,4,7], Ed DiBella[3,5], and Shireen Elhabian[1,2]

[1] Scientific Computing and Imaging Institute, University of Utah, SLC, UT
[2] Kahlert School of Computing, University of Utah, SLC, UT
[3] Department of Biomedical Engineering, University of Utah, SLC, UT
[4] Nora Eccles Harrison Cardiovascular Research and Training Institute, University of Utah, SLC, UT
[5] Department of Radiology and Imaging Sciences, University of Utah, SLC, UT
[6] Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD
[7] Division of Cardiology, University of Utah, SLC, UT

arefeen.sultan@utah.edu

Abstract. High-quality Late Gadolinium Enhancement (LGE) MRI can be helpful for atrial fibrillation management, yet scan quality is frequently compromised by patient motion, irregular breathing, and suboptimal image acquisition timing. While Multiple Instance Learning (MIL) has emerged as a powerful tool for automated quality assessment under weak supervision, current state-of-the-art methods map localized visual evidence to a single, opaque global feature vector. This *black box* approach fails to provide actionable feedback on specific failure modes, obscuring whether a scan degrades due to motion blur, inadequate contrast, or a lack of anatomical context. In this paper, we propose Adversarial Concept-MIL (AC-MIL), a weakly-supervised framework that decomposes global image quality into clinically defined radiological concepts using only volume-level supervision. To capture latent quality variations without entangling predefined concepts, our framework incorporates an unsupervised residual branch guided by an adversarial erasure mechanism to strictly prevent information leakage. Furthermore, we introduce a spatial diversity constraint that penalizes overlap between distinct concept attention maps, ensuring localized and interpretable feature extraction. Extensive experiments on a clinical dataset of atrial LGE-MRI volumes demonstrate that AC-MIL successfully opens the MIL *black box*, providing highly localized spatial concept maps that specify causes of non-diagnostic scans. Crucially, our framework achieves this deep clinical transparency while maintaining highly competitive ordinal grading performance against existing baselines. The code is available at: https://github.com/arf111/AC-MIL

Keywords: Image Quality Assessment · Concept Bottleneck Models · Multiple Instance Learning · Weak Supervision · Explainable AI.

1 Introduction

Atrial fibrillation (AF) is the most common form of cardiac arrhythmia, affecting an estimated 3 to 5 million individuals in the U.S., with projections indicating that this number could surpass 12 million by 2030 [3]. A key factor contributing to the development and recurrence of AF is atrial fibrosis, a structural remodeling of the heart characterized by the formation of fibrotic tissue that disrupts normal electrical conduction [5,14]. Catheter ablation, a widely adopted AF treatment, seeks to isolate the abnormal electrical signals. However, accurate fibrosis detection and quantification can help to estimate the success of ablation strategies. Despite advancements in ablation techniques, recurrence rates of AF remain high, sometimes exceeding 40% within 18 months following the procedure [19]. These challenges motivate improved fibrosis-assessment tools to guide treatment selection and reduce recurrence.

High resolution 3D Late Gadolinium Enhancement (LGE) MRI is used for detecting atrial fibrosis and scarring, playing a crucial role in generating patient-specific models to plan ablation procedures [15,1]. However, the diagnostic utility of LGE MRI is highly dependent upon image quality. Patient-related issues, such as movement and inconsistent breathing, can introduce motion artifacts. Additionally, suboptimal pulse sequence parameters and inadequate contrast administration may further degrade image clarity [8,6,17]. Manual evaluation of image quality is labor-intensive and challenging to integrate into fast-paced clinical workflows. While deep learning offers a pathway to automate image quality assessment (IQA), traditional supervised approaches require exhaustive pixel-level or slice-level artifact annotations, which are costly and impractical to obtain for large-scale 3D LGE MRI datasets. To address this annotation bottleneck, Multiple Instance Learning (MIL) [10] has emerged as a promising framework for learning under weak supervision. Specifically, the recent Sultan et al. [18] framework established the state-of-the-art for LGE-MRI quality assessment by employing a hierarchical architecture to predict volume-level diagnostic utility without explicit slice-level annotations.

However, standard MIL approaches, including recent hierarchical models, suffer from an inherent feature entanglement problem [18,10,12,20]. By mapping localized visual evidence to a single global feature vector, they operate as opaque black boxes. While this yields high accuracy, it obscures clinical reasoning. Furthermore, it leaves models susceptible to shortcut learning, where spurious visual cues artificially inflate performance instead of genuine anatomy. In clinical workflows, identifying that a scan is non-diagnostic is insufficient; technicians must know why: whether the failure is due to motion blur, inadequate contrast, or anatomical context. Concept Bottleneck Models (CBMs) [11] attempt to address this by aligning latent dimensions with human-interpretable concepts. However, naive CBMs applied to medical images frequently suffer from spatial overlap and information leakage, where independent concepts rely on the same entangled visual textures [4,9].

To bridge this gap, we propose AC-MIL (Adversarial Concept Multiple Instance Learning), a generalized, interpretable MIL framework that is domain-

agnostic and applicable to any weakly-supervised medical imaging task where clinical concepts can be defined. To overcome the inherent limitations of standard architectures, we present a novel framework that successfully integrates concept bottlenecks into a weakly-supervised MIL setting. Our contributions are as follows:

- *Disentangled Concept MIL Architecture:* We introduce a novel MIL formulation that explicitly decomposes the latent feature space into clinically meaningful concepts, transforming the opaque MIL pooling process into an interpretable clinical checklist.
- *Adversarially-Regularized Completeness:* We introduce an unsupervised residual branch to model latent quality variations not described by the predefined concepts, utilizing an adversarial erasure mechanism to strictly prevent information leakage.
- *Spatial Attention Diversity (SAD):* We propose a spatial diversity objective that penalizes overlap between the attention maps of distinct concepts to ensure localized and strictly interpretable feature extraction.

2 Method

AC-MIL consists of two hierarchical tiers: (1) slice-level disentanglement into concept-specific subspaces and (2) volume-level aggregation under asymmetric gradient control. The framework is optimized via a multi-objective loss enforcing concept supervision, adversarial orthogonality, and spatial diversity.

2.1 Clinical Quality Concepts and Rating Scale

The overall image quality for fibrosis assessment is evaluated on a 4-point ordinal scale. We decompose this assessment into three predefined, weakly-supervised volume-level concepts: *Sharpness*, evaluating edge definition and chamber blurring; *Myocardium Nulling*, assessing left ventricular intensity relative to the blood pool; and *Aorta and Valve Enhancement*, evaluates the visibility and contrast of the aorta and valve leaflet walls. To capture the remaining latent variations critical for the final quality assessment, we introduce a fourth *Unsupervised Residual* concept.

2.2 Problem Formulation and Architecture

We represent each 3D MRI volume as a nested hierarchy of pseudo-bags. Using segmentation masks, we restrict processing to the patient-specific number of axial slices (M) containing the left atrium. Let $\mathcal{V}$ be a bag of M slices $\{S_1, \ldots, S_M\}$, where $S_m \in \mathbb{R}^{H \times W}$. To accommodate the variable M across patients during training, we randomly sample a fixed-size sub-bag of $N \leq M$ slices. Each selected slice is further divided into a sub-bag of K random patches $\{x_1, \ldots, x_K\}$, with $x_k \in \mathbb{R}^{p \times p}$.

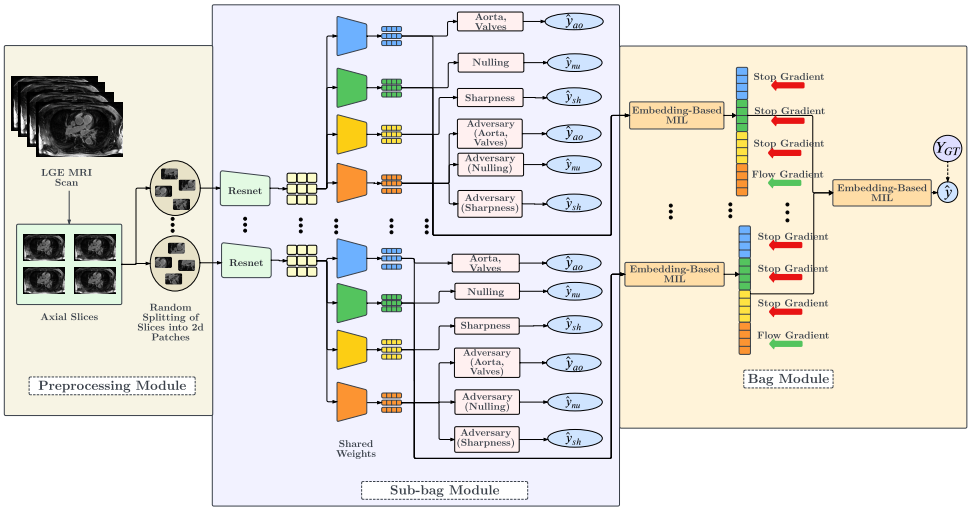


Fig. 1. Architecture of the proposed AC-MIL framework. 3D volumes are processed as a hierarchy of 2D slices and patches. Tier 1 (Sub-Bag Module) projects shared patch embeddings into three predefined clinical concept subspaces and one unsupervised branch (regularized by adversarial heads), using concept-specific attention for slice-level aggregation. Tier 2 (Bag Module) fuses these vectors via Attention-MIL to predict the final QA score.

2.3 Tier 1: Disentangled Sub-Bag Module

The first tier aims to decompose the entangled visual signals of a slice into our four distinct latent subspaces: Sharpness (Z_{sh}), Myocardium Nulling (Z_{nu}), Aorta/Valves (Z_{ao}), and the Unsupervised Residual (Z_{un}).

Concept-Specific Embedding. Each patch x_k is passed through a shared ResNet feature extractor $\Phi(\cdot)$ to obtain a feature embedding $h_k \in \mathbb{R}^d$, which is subsequently projected into 4 equal-dimensional concept subspaces via dedicated 2-layer fully connected projection heads $P^{(c)}$, yielding $h_k^{(c)} = P^{(c)}(h_k) \in R^L$ for $c \in \{sh, nu, ao, un\}$.

Concept-Specific Attention. To aggregate patches, we employ independent attention mechanisms for each subspace, enabling selective focus on concept-specific regions. The slice-level vector is $Z^{(c)} = \sum_{k=1}^{K} \alpha_k^{(c)} h_k^{(c)}$, where attention weights $\alpha_k^{(c)}$ are computed via a concept-specific gated attention mechanism followed by softmax normalization across the K patches.

2.4 Tier 2: Volume-Level Aggregation and Task Optimization

The slice-level representations are fused into a comprehensive vector V_m. To prevent the downstream task loss from forcing predefined concepts to illicitly encode final task information, known as Concept Leakage [9], we apply an asymmetric

stop-gradient, $\text{sg}(\cdot)$, during fusion. The slice-level representations are concatenated as $V_m = [\text{sg}(Z_{sh}), \text{sg}(Z_{nu}), \text{sg}(Z_{ao}), Z_{un}]$, where the stop-gradient operator prevents task-level gradients from propagating into the predefined concept subspaces. Consequently, these branches are optimized solely via their respective supervised concept losses, while the residual branch remains the only pathway through which task gradients can flow. Gradients from $\mathcal{L}_{task}$ still update $\Phi(\cdot)$ via the residual pathway, but cannot directly modify the predefined concept projections. A secondary Attention-MIL module then processes the sequence $V_{1...M}$ to compute slice-level weights β_m and aggregate them into a global volume representation $V_{vol} = \sum_{m=1}^{M} \beta_m V_m$.

A final classifier F_{task} processes V_{vol} to produce the volume-level Quality Assessment (QA) score. The primary objective, $\mathcal{L}_{task}$, optimizes this final prediction using the CORN loss [16] against the global ground-truth volume label y_{vol}:

$$\mathcal{L}_{task} = \mathcal{L}_{CORN}(F_{task}(V_{vol}), y_{vol}) \tag{1}$$

2.5 Learning via Multi-Objective Optimization

To ensure the clinical relevance of these subspaces, we optimize a joint loss function:

$$\mathcal{L}_{total} = \mathcal{L}_{task} + \lambda_{CBM}\mathcal{L}_{CBM} + \lambda_{adv}\mathcal{L}_{adv} + \lambda_{SAD}\mathcal{L}_{SAD} \tag{2}$$

Concept Supervision ($\mathcal{L}_{CBM}$): We apply CORN losses [16] to the predefined concept representations Z_{sh}, Z_{nu}, Z_{ao}. Due to the asymmetric stop-gradient, task-level gradients from $\mathcal{L}_{task}$ do not propagate into these branches. Consequently, their projections are optimized exclusively through $\mathcal{L}_{CBM}$, preventing direct task-driven concept leakage.

Adversarial Concept Erasure ($\mathcal{L}_{adv}$): To prevent the unconstrained residual branch Z_{un} from trivially relearning known semantics to minimize $\mathcal{L}_{task}$, we enforce feature orthogonality. Adversaries D_c attempt to predict clinical concept grades from the unsupervised vector, formulated as a minimax game via a Gradient Reversal Layer (GRL)[7]:

$$\mathcal{L}_{adv} = \sum_{c \in \{sh, nu, ao\}} \mathcal{L}_{CORN}(D_c(Z_{un}), y_c) \tag{3}$$

During the backward pass, the GRL reverses the gradients by a factor of $-\lambda_{adv}$ (i.e., $\frac{\partial GRL(x)}{\partial x} = -\lambda_{adv}\mathbf{I}$). This penalizes the feature extractor $\Phi(\cdot)$ if adversaries succeed, ensuring Z_{un} remains statistically independent from predefined concepts.

Spatial Attention Diversity ($\mathcal{L}_{SAD}$): We introduce a diversity constraint to prevent the model from collapsing multiple independent concepts onto the same anatomical patches. Crucially, global artifacts such as Sharpness affect the entire image and are therefore excluded from this penalty. We restrict the constraint

to localized concepts by defining the set $\mathcal{C}_{loc} = \{nu, ao, un\}$. We then minimize the cosine similarity between the spatial attention maps of these specific pairs:

$$\mathcal{L}_{SAD} = \sum_{\substack{c_i, c_j \in \mathcal{C}_{loc} \\ c_i \neq c_j}} \frac{\langle \alpha^{(c_i)}, \alpha^{(c_j)} \rangle}{\|\alpha^{(c_i)}\|_2 \|\alpha^{(c_j)}\|_2} \tag{4}$$

Minimizing this overlap forces the Aorta, Nulling, and Unsupervised heads to attend respectively to the vessel, the myocardium, and residual quality cues, guaranteeing spatial disentanglement.

3 Experiments & Results

3.1 Dataset

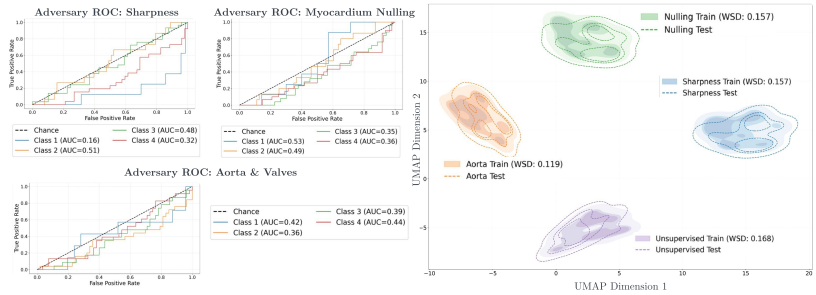
We evaluated our framework on 775 LGE MRI scans with acquisition resolution $1.25 \times 1.25 \times 2.5\ mm^3$, acquired at a single institution under IRB approval with informed consent or a documented waiver for retrospective use. Data was partitioned via stratified 10-fold cross-validation (10% test per fold; 10% of the remaining training portion held out for validation/early stopping). Original expert quality ratings on a 1–5 scale were collapsed into a 4-point ordinal scale to mitigate class imbalance, affecting 7% labeled samples that originally received a rating of 5. Global quality grades and the three concept-level grades (Sharpness, Myocardium Nulling, Aorta/Valve Enhancement) were each assigned manually by expert readers at the volume level, independent of one another, constituting our weak (volume-level only) supervision; no slice- or patch-level annotations were used. On a 50-case subset scored by 5 raters, inter-rater agreement (ICC) was 0.74 for overall quality and 0.62–0.72 across concept categories. We evaluate ordinal grading using Quadratic Weighted Kappa (QWK) and Average Mean Absolute Error (AMAE).

3.2 Implementation Details

During training, we sample sub-bags of $N = 8$ slices and $K = 80$ patches (64×64) per volume based on validation performance, whereas validation and testing employ dense sliding-window inference (50% patch overlap) across all M available slices. Models are optimized via AdamW [13] (learning rate 1×10^{-4}, weight decay 1×10^{-3}) with a 100-epoch early stopping patience on validation QWK. Tuned on the validation set to balance ordinal grading accuracy with spatial disentanglement, the loss weights in Eq. 2 are set to $\lambda_{CBM} = 1.0$, $\lambda_{adv} = 0.5$, and $\lambda_{SAD} = 0.1$.

3.3 Qualitative Disentanglement & Interpretability

Feature-Space Orthogonality and Erasure: To validate the adversarial erasure mechanism, Figure 2a presents the ROC curves of adversaries attempting



(a) AC-MIL (ROC of Unsupervised Branch)

(b) AC-MIL (UMAP)

Fig. 2. Feature-space orthogonality and adversarial erasure. (a) ROC curves of adversaries attempting to predict predefined concepts from the unsupervised branch. (b) UMAP projection of the learned feature spaces, with annotated Wasserstein Distances between training and test distributions.

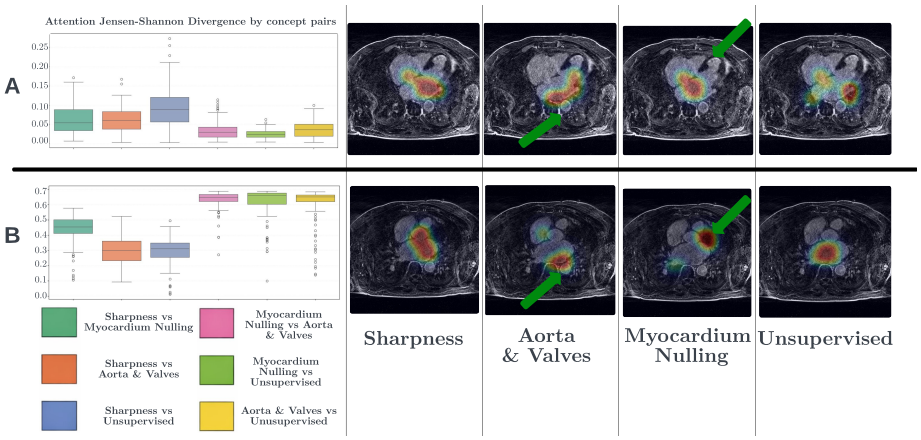
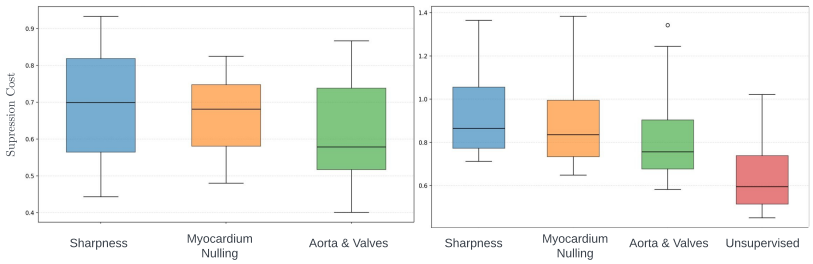


Fig. 3. Impact of the Spatial Attention Diversity ($\mathcal{L}_{SAD}$). (A) Without $\mathcal{L}_{SAD}$, showing Jensen-Shannon Divergence (JSD) boxplots for concept pairs (left) and entangled spatial attention maps (right). (B) With $\mathcal{L}_{SAD}$ applied, showing increased JSD and attention precisely localized on target anatomies (green arrows).

to predict predefined concepts from the unsupervised branch. The adversary ROC curves cluster around chance-level performance, indicating that $\mathcal{L}_{adv}$ effectively suppresses recoverable clinical concept information from the residual representation. Consequently, as visualized in the UMAP projection (Fig. 2b), the unsupervised representations do not entangle with existing features. Instead, they are forced to form a strictly distinct, non-overlapping fourth cluster alongside the predefined clinical concepts.

Spatial Disentanglement and Attention Diversity: To evaluate feature localization, we computed the Jensen-Shannon Divergence between the spatial attention distributions of distinct concept pairs. Without $\mathcal{L}_{SAD}$, JSD is low

(Fig. 3A), with Aorta and Nulling attention collapsing onto entangled, overlapping regions. This overlap allows the network to artificially inflate performance via shortcut learning, exploiting a single proxy variable rather than evaluating distinct clinical criteria. Applying $\mathcal{L}_{SAD}$ forces spatial specialization (Fig. 3B): predefined concepts decouple to target distinct anatomies (the aortic vessel and LV myocardium), while the unsupervised branch diverges to explore novel latent variations.



(a) Without Unsupervised (b) With Unsupervised

Fig. 4. Feature importance via adversarial attack. Box plots display the minimum L_2 perturbation cost required to degrade a high-quality scan for models trained (a) without and (b) with the unsupervised residual branch.

Clinical Relevance via Adversarial Attack: We evaluated feature importance via the minimum adversarial perturbation cost[2] needed to degrade a high-quality scan (Fig 4). The baseline model distributes sensitivity across predefined concepts, whereas full AC-MIL is sensitive to the unsupervised residual. This is clinically justified: aorta/myocardium concepts establish a baseline but cannot capture every failure mode. The unsupervised branch acts as a safety net, capturing complex anomalies a standard checklist would otherwise miss.

3.4 Quantitative Results and Ablation Study

We benchmarked AC-MIL against SOTA methods [10,18] and evaluated our components via an ablation study (Table 1). Our baseline AC-MIL variant establishes strong performance; however, Fig. 3 shows this peak score relies heavily on the aforementioned spatial entanglement. Applying $\mathcal{L}_{SAD}$ corrects this overlap, and the resulting marginal drop in the full model is a necessary trade-off to guarantee the network's reasoning aligns with true radiological anatomy.

Method	CBM	SAD	Adv	QWK ↑	AMAE ↓
ABMIL[10]				$0.60 \pm 0.03[0.56, 0.64]$	$0.59 \pm 0.01[0.56, 0.63]$
HAMIL-QA[18]				$0.63 \pm 0.01[0.60, 0.65]$	$0.57 \pm 0.02[0.54, 0.61]$
AC-MIL	✓			$0.65 \pm 0.01[0.63, 0.67]$	$0.56 \pm 0.01[0.53, 0.58]$
AC-MIL	✓	✓		$0.64 \pm 0.02[0.61, 0.67]$	$0.57 \pm 0.02[0.53, 0.61]$
AC-MIL	✓		✓	$\mathbf{0.68 \pm 0.01[0.65, 0.70]}$	$\mathbf{0.53 \pm 0.01[0.50, 0.56]}$
AC-MIL(Full)	✓	✓	✓	$\mathbf{0.66 \pm 0.01[0.64, 0.67]}$	$\mathbf{0.55 \pm 0.01[0.52, 0.57]}$

Table 1. Ablation study on the components of our proposed framework on 10 folds with 95% confidence interval.

4 Conclusion

In this paper, we introduced AC-MIL, a novel weakly-supervised framework for the interpretable quality assessment of atrial LGE-MRI scans. By preventing information leakage and penalizing shortcut learning, our approach successfully disentangles global quality predictions into human-interpretable radiological concepts. A key limitation is that evaluation is restricted to a single-center, in-house dataset, as no public benchmark currently provides both ordinal QA and concept-level labels for atrial LGE-MRI; multi-center validation remains important future work.

Acknowledgments. Research reported in this publication was supported by the National Heart, Lung, and Blood Institute and the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under award numbers R01HL162353 and R50EB038027. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

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

References

1. Caixal, G., Alarcón, F., Althoff, T.F., Nuñez-Garcia, M., Benito, E.M., Borràs, R., Perea, R.J., Prat-González, S., Garre, P., Soto-Iglesias, D., Gunturitz, C., Cozzari, J., Linhart, M., Tolosana, J.M., Arbelo, E., Roca-Luque, I., Sitges, M., Guasch, E., Mont, L.: Accuracy of left atrial fibrosis detection with cardiac magnetic resonance: correlation of late gadolinium enhancement with endocardial voltage and conduction velocity. Europace **23**(3), 380–388 (Mar 2021). https://doi.org/10.1093/europace/euaa313
2. Carlini, N., Wagner, D.: Towards evaluating the robustness of neural networks. In: 2017 ieee symposium on security and privacy (sp). pp. 39–57. Ieee (2017)
3. Colilla, S., Crow, A., Petkun, W., Singer, D.E., Simon, T., Liu, X.: Estimates of Current and Future Incidence and Prevalence of Atrial Fibrillation in the U.S. Adult Population. The American Journal of Cardiology **112**(8), 1142–1147 (Oct 2013). https://doi.org/10.1016/j.amjcard.2013.05.063
4. Corbetta, V., Dijkstra, F.S., Beets-Tan, R., Kervadec, H., Wickstrøm, K., Silva, W.: In-hoc concept representations to regularise deep learning in medical imaging.

In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 7312–7321 (2025)
5. ElMaghawry, M., Romeih, S.: DECAAF: Emphasizing the importance of MRI in AF ablation. Glob Cardiol Sci Pract **2015**, 8 (Mar 2015). https://doi.org/10.5339/gcsp.2015.8
6. Flett, A.S., Hasleton, J., Cook, C., Hausenloy, D., Quarta, G., Ariti, C., Muthurangu, V., Moon, J.C.: Evaluation of techniques for the quantification of myocardial scar of differing etiology using cardiac magnetic resonance. JACC: cardiovascular imaging **4**(2), 150–156 (2011)
7. Ganin, Y., Lempitsky, V.: Unsupervised domain adaptation by backpropagation. In: International conference on machine learning. pp. 1180–1189. PMLR (2015)
8. Gräni, C., Eichhorn, C., Bière, L., Kaneko, K., Murthy, V.L., Agarwal, V., Aghayev, A., Steigner, M., Blankstein, R., Jerosch-Herold, M., et al.: Comparison of myocardial fibrosis quantification methods by cardiovascular magnetic resonance imaging for risk stratification of patients with suspected myocarditis. Journal of Cardiovascular Magnetic Resonance **21**, 1–11 (2019)
9. Havasi, M., Parbhoo, S., Doshi-Velez, F.: Addressing leakage in concept bottleneck models. Advances in Neural Information Processing Systems **35**, 23386–23397 (2022)
10. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International conference on machine learning. pp. 2127–2136. PMLR (2018)
11. Koh, P.W., Nguyen, T., Tang, Y.S., Mussmann, S., Pierson, E., Kim, B., Liang, P.: Concept bottleneck models. In: International conference on machine learning. pp. 5338–5348. PMLR (2020)
12. Li, B., Li, Y., Eliceiri, K.W.: Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 14318–14328 (2021)
13. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
14. Marrouche, N.F., Wilber, D., Hindricks, G., Jais, P., Akoum, N., Marchlinski, F., Kholmovski, E., Burgon, N., Hu, N., Mont, L., Deneke, T., Duytschaever, M., Neumann, T., Mansour, M., Mahnkopf, C., Herweg, B., Daoud, E., Wissner, E., Bansmann, P., Brachmann, J.: Association of Atrial Tissue Fibrosis Identified by Delayed Enhancement MRI and Atrial Fibrillation Catheter Ablation: The DECAAF Study. JAMA **311**(5), 498–506 (Feb 2014). https://doi.org/10.1001/jama.2014.3
15. Oakes, R.S., Badger, T.J., Kholmovski, E.G., Akoum, N., Burgon, N.S., Fish, E.N., Blauer, J.J.E., Rao, S.N., DiBella, E.V.R., Segerson, N.M., Daccarett, M., Windfelder, J., McGann, C.J., Parker, D., MacLeod, R.S., Marrouche, N.F.: Detection and quantification of left atrial structural remodeling with delayed-enhancement magnetic resonance imaging in patients with atrial fibrillation. Circulation **119**(13), 1758–1767 (Apr 2009). https://doi.org/10.1161/CIRCULATIONAHA.108.811877
16. Shi, X., Cao, W., Raschka, S.: Deep neural networks for rank-consistent ordinal regression based on conditional probabilities. Pattern Analysis and Applications **26**(3), 941–955 (2023)
17. Spiewak, M., Malek, L.A., Misko, J., Chojnowska, L., Milosz, B., Klopotowski, M., Petryka, J., Dabrowski, M., Kepka, C., Ruzyllo, W.: Comparison of different quantification methods of late gadolinium enhancement in patients with hypertrophic cardiomyopathy. European journal of radiology **74**(3), e149–e153 (2010)

18. Sultan, K.A., Hisham, M.H.H., Orkild, B., Morris, A., Kholmovski, E., Bieging, E., Kwan, E., Ranjan, R., DiBella, E., Elhabian, S.: Hamil-qa: Hierarchical approach to multiple instance learning for atrial lge mri quality assessment. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 275–284. Springer (2024)
19. Verma, A., Jiang, C.y., Betts, T.R., Chen, J., Deisenhofer, I., Mantovan, R., Macle, L., Morillo, C.A., Haverkamp, W., Weerasooriya, R., Albenque, J.P., Nardi, S., Menardi, E., Novak, P., Sanders, P.: Approaches to Catheter Ablation for Persistent Atrial Fibrillation. New England Journal of Medicine **372**(19), 1812–1822 (May 2015). https://doi.org/10.1056/NEJMoa1408288, publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJMoa1408288
20. Zhang, H., Meng, Y., Zhao, Y., Qiao, Y., Yang, X., Coupland, S.E., Zheng, Y.: Dtfd-mil: Double-tier feature distillation multiple instance learning for histopathology whole slide image classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18802–18812 (2022)