# HAMIL-QA: Hierarchical Approach to Multiple Instance Learning for Atrial LGE MRI Quality Assessment

K M Arefeen Sultan[1,2], Md Hasibul Husain Hisham[1,2], Benjamin Orkild[1,3,4], Alan Morris[1], Eugene Kholmovski[5,6], Erik Bieging[5,7], 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, Baltimore, MD
[7] Division of Cardiology, University of Utah, SLC, UT

**Abstract.** The accurate evaluation of left atrial fibrosis via high-quality 3D Late Gadolinium Enhancement (LGE) MRI is crucial for atrial fibrillation management but is hindered by factors like patient movement and imaging variability. The pursuit of automated LGE MRI quality assessment is critical for enhancing diagnostic accuracy, standardizing evaluations, and improving patient outcomes. The deep learning models aimed at automating this process face significant challenges due to the scarcity of expert annotations, high computational costs, and the need to capture subtle diagnostic details in highly variable images. This study introduces HAMIL-QA, a multiple instance learning (MIL) framework, designed to overcome these obstacles. HAMIL-QA employs a hierarchical bag and sub-bag structure that allows for targeted analysis within sub-bags and aggregates insights at the volume level. This hierarchical MIL approach reduces reliance on extensive annotations, lessens computational load, and ensures clinically relevant quality predictions by focusing on diagnostically critical image features. Our experiments show that HAMIL-QA surpasses existing MIL methods and traditional supervised approaches in accuracy, AUROC, and F1-Score on an LGE MRI scan dataset, demonstrating its potential as a scalable solution for LGE MRI quality assessment automation. The code is available at: https://github.com/arf111/HAMIL-QA

**Keywords:** Image Quality Assessment · Weak Supervision · Multiple Instance Learning · Attention-based Models

## 1 Introduction

Atrial fibrillation (AF), the most prevalent type of cardiac arrhythmia in the U.S., currently affects between 3 and 5 million individuals [3]. Projections suggest

this number could surge to over 12 million by 2030 [3]. Research has established a significant connection between atrial fibrosis and the onset and recurrence of AF post-treatment [4,10]. Catheter ablation, a widely adopted approach for treating AF, focuses on eradicating fibrotic tissues in the heart responsible for erratic electrical impulses by forming precise lesions or scars in these areas. This underscores the importance of accurately measuring fibrosis to effectively steer the ablation process. Despite its popularity, catheter ablation's efficacy is limited, with a recurrence rate of AF exceeding 40% within 18 months post-procedure [15]. This high recurrence rate highlights the critical need to address the limitations of current AF treatments.

Late Gadolinium Enhancement (LGE) MRI is widely utilized to quantify myocardial fibrosis and scarring. It is instrumental in assessing AF patients before catheter ablation, providing detailed insights into the atrial structure and fibrosis distribution. The geometry and fibrosis patterns identified through LGE MRI are essential for planning ablation procedures and generating patient-specific models [11,2]. However, the quality of LGE MRI images can vary significantly, influenced by aspects such as noise, patient mobility, inconsistent breathing patterns, and suboptimal tuning of pulse sequence parameters; which can impact diagnostic precision [7,6,13].

Automating the quality assessment (QA) of LGE MRI images is critically important for clinical practices, offering to improve diagnostic accuracy, increase procedural efficiency, standardize assessments, and enhance patient outcomes. By ensuring high-quality scans for fibrosis quantification, this automation directly contributes to refining ablation strategies and guiding treatments more effectively. However, manual evaluation of image quality is labor-intensive and susceptible to errors, making it unsuitable for widespread application and challenging its integration into clinical practice. Automating QA can be approached naively by training a deep network that would predict a quality score from a 3D LGE volume. However, this straightforward approach faces several challenges. Firstly, the constraint of limited annotations, predominantly expert-driven, significantly hinders the compilation of large, annotated datasets essential for such conventional deep learning paradigms. Secondly, the computational and memory requirements necessary for processing entire 3D volumes for image-level prediction pose a significant scalability challenge. Moreover, the variability in LGE MRI images due to patient anatomy differences and motion artifacts complicates the development of a generalizable model. Lastly, ensuring that the network's predictions are clinically relevant is a challenge, as the model must discern subtle quality nuances impacting diagnostic outcomes.

The quality assessment of LGE MRI scans inherently requires a weakly supervised learning approach, as typically only image-level labels are available. Multiple Instance Learning (MIL), by design, is well-suited to address this problem. MIL conceptualizes each 3D volume as a collection of instances (or patches), relying solely on the volume-level class label. The primary objective is to train a model capable of predicting the label for a group of instances, or "bag". In the context of our work, this entails the evaluation of image quality for fibrosis de-

tection in LGE MRI scans. Each scan, considered as a *bag*, comprises numerous instances, represented by hundreds of patches extracted from the scan. A bag is classified as positive if it contains at least one instance of diagnostic fibrotic tissue; otherwise, it is deemed a non-diagnostic scan. The choice of MIL as the foundational framework is justified by the nature of LGE MRI images, which does not guarantee the presence of diagnostic fibrotic tissue in every instance within a bag. Moreover, MIL optimizes computational efficiency by processing image patches rather than volumes and learns to focus on the most informative patches, thus reducing the overall processing load. This targeted approach aids in constructing models that are both generalizable and robust, capable of learning effectively from a limited number of annotated volumes in contrast to training networks that estimate volume-level labels from the full 3D volume. Furthermore, the MIL framework identifies the most diagnostic instances within a volume, aligning the learning process more closely with clinical relevance.

In this paper, we introduce HAMIL-QA, an MIL approach inspired by the cognitive processes employed by radiologists when assessing the diagnostic quality of LGE MRI images for the quantification of fibrosis. Although these processes may vary among radiologists, we followed the approach used by the radiologists who scored the scans in our dataset. This approach ingeniously models the radiologist's mental strategy of evaluating scans, employing a hierarchical structure of bags and sub-bags that mimics the expert's method of systematically sweeping through the scan, slice by slice, to determine the overall quality label. In this framework, MIL is applied first at the sub-bag level, focusing on slices within each volume, and then at a higher bag level, integrating features learned from the sub-bag level to determine the final quality score for the entire volume. This tiered structure allows for nuanced analysis and interpretation of image data, enhancing the model's effectiveness and efficiency. Specifically, at the sub-bag level, the model learns to identify the most informative instances. This selective attention increases the model's ability to handle the inherent variability in LGE MRI images, as it learns to recognize and adapt to patterns across different patient anatomies and artifact influences within the sub-bags before integrating these insights at the bag level. Experimental findings reveal that HAMIL-QA surpasses both traditional fully supervised models and current MIL methodologies in achieving higher accuracy, AUROC, and F1-Score metrics on a limited labeled dataset of LGE MRI scans.

## 2 Related Works

In MRI image quality control, advancements have been made to enhance the diagnostic accuracy and consistency of MRI scans. For instance, Wang et al. proposed deep generative model to enhance the quality control process of cardiac MRI segmentation, demonstrating its effectiveness across various datasets [16]. Dormont et al. proposed a framework for the automatic quality control of 3D brain T1-weighted MRI for a large clinical data warehouse [1]. K. Sultan et al. proposed a two-stage deep learning model to automate the quality assessment

of LGE MRI images, crucial for evaluating left atrial fibrosis in patients with atrial fibrillation [14].

Within medical image analysis, MIL has emerged as a potent framework, given its efficacy for classification tasks in histopathology. Quellec et al. explored MIL's applicability across various medical image and video analysis tasks, demonstrating its potential to circumvent the need for detailed pixel-level annotations [12]. One notable contribution is from Ilse et al., who proposed an attention-based deep MIL, enhancing interpretability and performance in medical image analysis by focusing on relevant instances within a bag [8]. Furthermore, Shao et al. introduced a Dual-stream MIL Network that leverages whole slide image classification integrated with self-supervised contrastive learning, providing a nuanced approach to learning from unannotated regions within medical images effectively [9]. Additionally, the DTFD-MIL approach by Li et al. emphasizes a double-tier feature distillation within a MIL framework, addressing the complexities of histopathology image classification with refined feature representation and enhanced learning efficacy [17].

To the best of our knowledge, the confluence of MIL methodologies in LGE-MRI QA remains uninvestigated. Notably, the comprehensive review by Fatima et al. on MIL underscores its versatility across a spectrum of applications, from image retrieval to disease diagnosis, reinforcing the value of integrating such methodologies within LGE-MRI QA processes [5].

## 3   Method

Consider an LGE MRI volume $\mathbf{V} \in \mathbb{R}^{R \times C \times S}$, where $R$, $C$, and $S$ represent the number of rows, columns, and slices of the volume, respectively. The diagnostic quality of the LGE MRI volume is defined as the ground truth label $Y$, which categorizes the scan as either diagnostic or non-diagnostic. For notational simplicity and without loss of generality, we assume each volume comprises an equal number of $M$ sub-bags, represented as $\mathcal{X} = \{\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_M\}$, with each sub-bag $\mathcal{X}_m$ corresponding to an axial slice. A sub-bag inherits its label from its parent bag's label $Y_{GT}$. It is also assumed that each sub-bag contains a consistent number of $K$ instances, denoted by $\mathcal{X}_m = \{\mathbf{x}_m^1, \mathbf{x}_m^2, \ldots, \mathbf{x}_m^K\}$, where $\mathbf{x}_m^k \in \mathbb{R}^{r \times c}$ is a 2D image patch of $r-$rows and $c$-columns. We randomly sample 2D patches from each axial slice to construct a sub-bag, reflecting the inherent variability within the scan. It is important to note that the proposed approach does not inherently require that each volume has the same number of sub-bags or that each sub-bag contains an identical number of instances, allowing for flexibility in handling diverse LGE MRI data structures.

We apply a transformation function $\phi(.)$ on each instance (i.e., patch) $\mathbf{x}_m^k$ to obtain the instances embedding (i.e., feature descriptors) for each sub-bag, denoted as $\mathcal{H}_m = \{\mathbf{h}_m^1, \mathbf{h}_m^2, \ldots, \mathbf{h}_m^K\}$, where $\mathbf{h}_m^k = \phi(\mathbf{x}_m^k) \in \mathbb{R}^{L \times 1}$. We use ResNet10 to parameterize the transformation function $\phi(.)$. Then, we take a dual-stream approach on the features to perform both embedding-based MIL and slice feature distillation in our sub-bag module. In the first stream which is
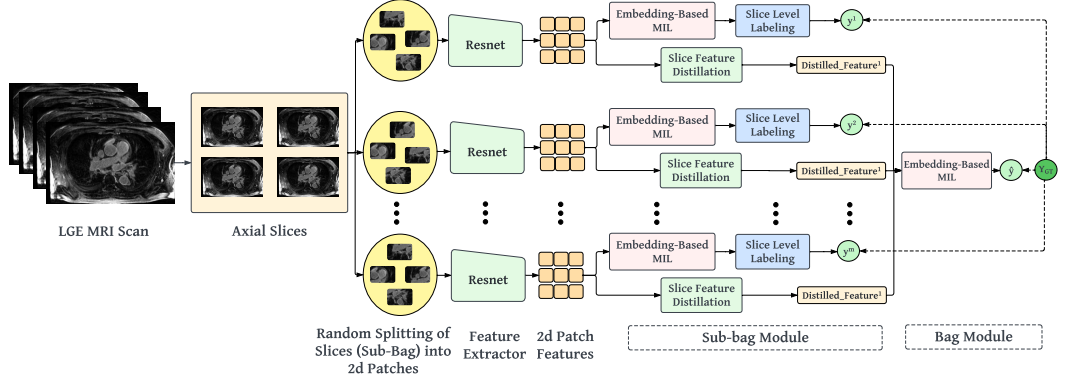
Fig. 1: Overview of our proposed model. For illustrative purposes, we select M axial slices (with 4 depicted as an example) at random from an LGE MRI scan, treating each slice as an individual sub-bag. We then extract random cropped patches from each of these slices. These sub-bags are initially processed by the sub-bag module. Subsequently, the output from sub-bag module is used to generate feature vectors, which are then input into the bag module. It is important to note that the ground truth label for the bags remains consistent across both sub-bag module and bag module during the training phase.

the embedding-based MIL as denoted in Figure 1, we get the slice or sub-bag level prediction label, $\hat{y_m} = \rho_{\text{sub-bag}}(\mathcal{H}_m)$. This involves computing an attention-weighted sum of features:

$$\hat{y}_m = \rho_{\text{sub-bag}}(\mathcal{H}_m) = \sigma\left(\mathbf{w}_m \sum_{k=1}^{K} a_m^k \mathbf{h}_m^k\right), \qquad (1)$$

where $a_m^k$ denotes the attention weight for the $k$-th instance in the $m-$th sub-bag $\mathcal{H}_m$, $\mathbf{w}_m \in \mathbb{R}^{L\times 1}$ denotes a weight vector for binary classification, and $\sigma$ represents a sigmoid activation mapping the aggregated signal to a prediction. Similar to [8], we calculate the attention as:

$$a_m^k = \frac{\exp\left\{\mathbf{w}^\top \tanh(\mathbf{V}\mathbf{h}_m^{k\top})\right\}}{\sum_{j=1}^{K} \exp\left\{\mathbf{w}^\top \tanh(\mathbf{V}\mathbf{h}_m^{j\top})\right\}} \qquad (2)$$

Consider a dataset of $N-$volumes. The sub-bag module loss is then calculated as:

$$\mathcal{L}_{\text{sub-bag}} = -\frac{1}{MN} \sum_{n=1}^{N} \sum_{m=1}^{M} \left[Y_n^m \log(y_n^m) + (1 - Y_n^m) \log(1 - y_n^m)\right] \qquad (3)$$

On the other stream, slice feature distillation, we get the distilled features from transformation function $\phi(.)$ calculated as:

$$\bar{\mathbf{h}} = \frac{1}{M} \sum_{n=1}^{M} \mathbf{h}_n \in \mathbb{R}^L \tag{4}$$

In the subsequent bag module, we use an embedding based MIL from the distilled features $\bar{\mathbf{h}}$ to obtain the final LGE MRI scan level label, $\hat{y}$ by following the same procedure as equation 1 and 2. The bag module loss is calculated as:

$$\mathcal{L}_{\text{bag}} = -\frac{1}{N} \sum_{n=1}^{N} [Y_n \log(\hat{y}_n) + (1 - Y_n) \log(1 - \hat{y}_n)] \tag{5}$$

Overall optimization is then:

$$\theta^* = \arg\min_{\theta_1, \theta_2} (\mathcal{L}_{\text{sub-bag}} + \mathcal{L}_{\text{bag}}) \tag{6}$$

where $\theta_1$ and $\theta_2$ are the parameters of Sub-bag module and Bag module, respectively.

## 4    Results

### 4.1    Dataset

In this study, we employed a dataset comprising 424 LGE MRI scans, each labeled for the purpose of Quality Assessment (QA) and has Left Atrium segmentations. Following the acquisition protocol detailed in [10], scans were obtained with a fine resolution of $1.25 \times 1.25 \times 2.5\text{mm}^3$, captured roughly 15 minutes post the administration of gadolinium using a 3D ECG-gated and respiratory-navigated gradient echo inversion recovery sequence. Expert reviewers rated these scans on a scale from 1 to 5. These 424 scans have a class imbalance problem because most scans are in the 2 to 4 range. To address this problem, we have transformed the scores into two different labels: diagnostic and non-diagnostic. Scans with a score of $\geq 3$ are designated as diagnostic, denoted 1, while less than 3 is non-diagnostic and denoted as 0.

### 4.2    Data Preprocessing

Our dataset was splitted into training and testing sets following an 80:20 split. Subsequently, the training set was further partitioned into a secondary training set and a validation set, adhering to the same 80:20 ratio. For the proposed method, we processed each scan as a series of 2D axial slices, selectively utilizing those representing the Left Atrium based on available segmentation data. In contrast with other methodologies, as detailed in Table 1, we extracted a sub-volume inclusive of the Left Atrium from the 3D scans based on the available segmentations data. We expand this subvolume by 30 units along the axial plane only. Prior to network input, all data underwent a normalization process.

### 4.3    Experiments

During training, the model is validated using AUROC across all volumes. The performance is measured by Accuracy, AUROC, and F1-Score. The details of the training are given below:

We consider 3 models for our comparison. Fully supervised, Attention based MIL [8], and DTFD-MIL framework [17]. For the fully supervised model, we only consider a 3D ResNet10 model. The 3D images are passed through the network to predict binary score, i.e., non-diagnostic and diagnostic. For ABMIL and DTFD-MIL models, we randomly pick 3D patches from the 3D images. All of the networks were trained for 50 epochs using the Adam optimizer with a learning rate of $\eta = 0.0001$ and with an early stopping criteria of patience 8. A cosine annealing learning rate scheduler was used to reduce the learning rate throughout training. The results are shown in Table 1.
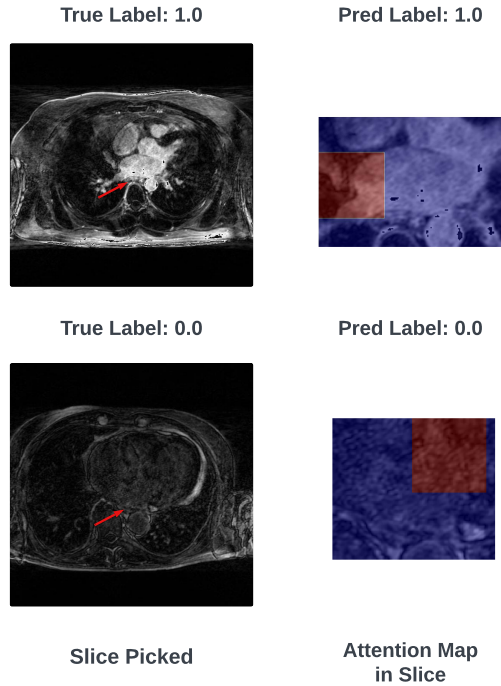


Fig. 2: Heatmap of 2 scans by original LGE MRI image and by attention map of our model, respectively. In the second column, a highlighted red square marks the patch receiving the highest attention weight, with an enlarged view provided for clarity. Additionally, a red arrow on the original MRI images indicates the left atrium's position.

| Method | Acc | AUROC | F1 |
|---|---|---|---|
| Fully Supervised | $0.545 \pm 0.062$ | $0.544 \pm 0.059$ | $0.419 \pm 0.210$ |
| Classic AB-MIL [8] | $0.643 \pm 0.028$ | $0.647 \pm 0.013$ | $0.433 \pm 0.093$ |
| DTFD-MIL (AFS) [17] | $0.604 \pm 0.075$ | $0.637 \pm 0.065$ | $0.242 \pm 0.242$ |
| HAMIL-QA | $\mathbf{0.682 \pm 0.030}$ | $\mathbf{0.700 \pm 0.009}$ | $\mathbf{0.596 \pm 0.084}$ |

Table 1: Results on our LGE MRI test set. For DTFD-MIL and our method (HAMIL-QA). the number of sub-bags is 6 and number of instances are 60. These numbers were determined by hyperparameter tuning. We further show the ablation experiments on these parameters in supplementary material. All of the experiments are run 3 times. The best ones are in bold.

For experiments, we considered different number of instances and pseudo bags to find the optimal performance. We also reported that our model is 700x and 89x more efficient in computation than fully supervised and two other models, respectively, since our model processes 2D image patches instead of full volume or 3D patches. The experiments are shown in the supplementary material.

To delve deeper into the efficacy of our approach, we visualized the model's attention mechanism by producing attention score heatmaps for two scans, as depicted in Figure 2. The visualizations reveal that our model appropriately allocates higher attention to the walls of the left atrium, which aligns with the regions commonly associated with a higher probability of fibrosis, indicating the model's capability to focus on clinically significant areas.

## 5    Conclusion

In conclusion, our study shows the development of a dual-module Multiple Instance Learning (MIL) framework is specifically designed to enhance the diagnostic quality assessment of Late Gadolinium Enhancement (LGE) Magnetic Resonance Imaging (MRI) scans. The introduction of a sub-bag concept and a double module mechanism effectively addresses the prevalent challenge of limited annotated datasets, significantly improving the model's performance metrics. The findings from this study demonstrate the framework's superior performance over traditional fully supervised and existing MIL methodologies. In summary, our study offers important contributions towards evaluating the quality of the left atrium in LGE MRI images, particularly when faced with a scarcity of labeled data.

## Acknowledgements

## Disclosure of Interests

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

## References

1. Bottani, S., Burgos, N., Maire, A., Wild, A., Ströer, S., Dormont, D., Colliot, O., Group, A.S., et al.: Automatic quality control of brain t1-weighted magnetic resonance images for a clinical data warehouse. Medical Image Analysis **75**, 102219 (2022)
2. 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
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. 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
5. Fatima, S., Ali, S., Kim, H.C.: A comprehensive review on multiple instance learning. Electronics **12**(20), 4323 (2023)
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. 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)
8. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: International conference on machine learning. pp. 2127–2136. PMLR (2018)
9. 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)
10. 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

11. 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
12. Quellec, G., Cazuguel, G., Cochener, B., Lamard, M.: Multiple-instance learning for medical image and video analysis. IEEE Reviews in Biomedical Engineering **10**, 213–234 (2017). https://doi.org/10.1109/RBME.2017.2651164
13. 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)
14. Sultan, K.A., Orkild, B., Morris, A., Kholmovski, E., Bieging, E., Kwan, E., Ranjan, R., DiBella, E., Elhabian, S.: Two-stage deep learning framework for quality assessment of left atrial late gadolinium enhanced mri images. In: International Workshop on Statistical Atlases and Computational Models of the Heart. pp. 230–239. Springer (2023)
15. 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
16. Wang, S., Tarroni, G., Qin, C., Mo, Y., Dai, C., Chen, C., Glocker, B., Guo, Y., Rueckert, D., Bai, W.: Deep generative model-based quality control for cardiac mri segmentation. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part IV 23. pp. 88–97. Springer (2020)
17. 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)