Robust Amyloidosis Subtype Classification via Multisequence CMR Fusion with Spatiotemporal Learning

Parker Martin[1] PARKER.MARTIN@OSUMC.EDU
Yifan Liu[1] LIU.11275@BUCKEYEMAIL.OSU.EDU
Karolina M. Zareba[1] KAROLINA.ZAREBA@OSUMC.EDU
Suzanne Smart[1] SUZANNE.SMART@OSUMC.EDU
Akash Goyal[1] AKASH.GOYAL@OSUMC.EDU
Orlando P. Simonetti[1] ORLANDO.SIMONETTI@OSUMC.EDU
Jeremy Slivnick[2] JSLIVNICK@BSD.UCHICAGO.EDU
Yuan Xue[1] YUAN.XUE@OSUMC.EDU
[1] *The Ohio State University, Columbus, OH, USA*
[2] *University of Chicago, Chicago, IL, USA*

Editors: Under Review for MIDL 2025

Abstract

Cardiac amyloidosis (CA) subtype classification remains a critical diagnostic challenge. We propose a multimodal deep learning framework that integrates cine, late gadolinium enhancement (LGE), and T1/T2 parametric cardiac MRI sequences to differentiate light chain (AL) and transthyretin (ATTR) amyloidosis. The model employs sequence-specific encoders and gated attention fusion, enabling robust performance even with missing input sequences. Evaluated on 123 patients with cross-validation, the xLSTM-based model achieved the highest AUC (0.8506), outperforming a Video Swin Transformer (VST) alternative. Grad-CAM visualizations highlight both cardiac and extracardiac regions, demonstrating interpretability and the potential for identifying systemic imaging biomarkers. These results support a clinically viable approach to non-invasive CA subtype diagnosis.

Keywords: Cardiac Amyloidosis, Cardiac MRI, Deep Learning, Multimodal Learning, Spatiotemporal Modeling, Model Interpretability

1. Introduction

Cardiac amyloidosis (CA) is a progressive infiltrative cardiomyopathy caused by extracellular deposition of misfolded proteins, most commonly light chain (AL) or transthyretin (ATTR) amyloid. Accurate subtype classification is essential for guiding therapy and prognosis (Maggialetti et al., 2024; Aus dem Siepen and Hansen, 2024), but current pathways often rely on invasive biopsy or specialized nuclear imaging (Dorbala et al., 2020). Cardiac MRI (CMR) offers a non-invasive and information-rich alternative (Fontana et al., 2015a), yet interpreting cine, late gadolinium enhancement (LGE), and T1/T2 mapping requires expert knowledge and remains a challenge (Fontana et al., 2015b; Banypersad et al., 2015; Zhao et al., 2016). We present a multimodal deep learning framework that combines these CMR sequences using sequence-specific encoders and gated attention fusion (Azam et al., 2022), with a focus on spatiotemporal learning from cine using xLSTM (Beck et al., 2024). The model supports missing sequences and provides interpretable Grad-CAM visualizations (Selvaraju et al., 2017), enabling robust and explainable CA subtype classification.

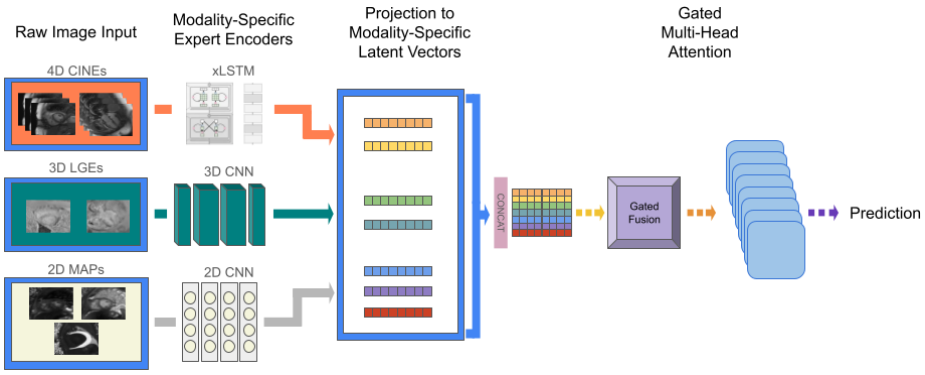

Figure 1: Model architecture illustrating sequence-specific encoders and gated multi-head attention fusion.

2. Methodology

Dataset and Preprocessing. We used cardiac MRI scans from 123 patients with biopsy-confirmed amyloidosis at The Ohio State University. Each exam included cine sequences (2D time-resolved), LGE (multi-slice 2D stacks), and T1/T2 parametric maps. While acquired as 2D or 3D in clinical imaging, we refer to their dimensionality based on model input structure. All images were resampled to 0.94 mm isotropic resolution. Intensity normalization used z-scoring for cine and LGE, and percentile scaling for parametric maps.
Model Architecture. Our framework (Fig. 1) employs sequence-specific encoders: a spatiotemporal network (xLSTM (Beck et al., 2024) or Video Swin Transformer (Liu et al., 2022)) for cine MRI, a 3D CNN for LGE, and a 2D CNN for parametric maps. Each encoder outputs a latent representation, which is dynamically fused using gated multi-head attention (Vaswani et al., 2017; Bahdanau et al., 2014). This design supports robustness to missing modalities while preserving temporal structure in cine sequences.
Training Strategy. We conducted 5-fold stratified cross-validation to compare model variants: (1) xLSTM vs. VST for cine encoding, and (2) with or without demographics (age, sex). Single-sequence baselines were also evaluated. Models were trained with cross-entropy loss, Adam optimizer (Kingma and Ba, 2015), dropout, and label smoothing. Hyperparameters were selected via Bayesian optimization using TPE (Bergstra et al., 2013).

3. Results

Model Comparison. Cross-validation results (Table 1) show that the spatiotemporal xLSTM achieved the highest AUC (0.8506 ± 0.0654) with lower computational cost. VST offered slightly lower accuracy (0.8346 ± 0.0410) at nearly double the parameter count and

Table 1: Cross-validation performance comparison across configurations.

Model Configuration	AUC	Inference Time (ms)	Min/Med/Max (ms)	# Params	Peak GPU (MB)
Spatiotemporal xLSTM	0.8506 ± 0.0654	65.42 ± 45.50	46.12 / 53.73 / 283.31	137M	71.67
Video Swin Transformer	0.8346 ± 0.0410	107.31 ± 36.63	89.88 / 97.31 / 309.29	276M	97.76
xLSTM + Demographics	0.8244 ± 0.0820	61.86 ± 44.57	45.06 / 50.26 / 275.59	137M	71.67
VST + Demographics	0.8154 ± 0.0656	110.79 ± 37.10	93.67 / 101.56 / 298.95	276M	97.76

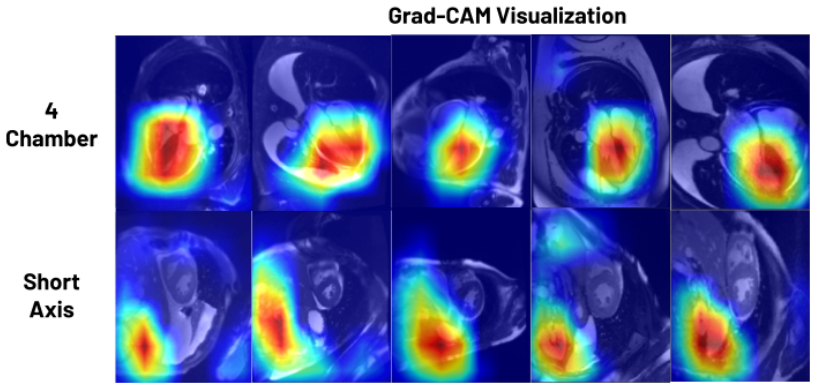

Figure 2: Grad-CAM visualization highlighting myocardial and extracardiac regions (e.g., kidneys), suggesting systemic features relevant to subtype differentiation.

memory usage. Including demographic data modestly reduced performance in both models (AUCs 0.8244 and 0.8154), likely due to overfitting given the limited dataset size.

Single-Sequence Instability. Models trained on individual sequences failed to converge reliably across folds, regardless of architecture. This instability reinforces the value of multimodal fusion, suggesting that complementary information across sequences is essential for subtype discrimination.

Interpretability and Systemic Insights. Fig. 2 displays Grad-CAM visualizations showing myocardial and extracardiac activations. Attention weights were consistently balanced across modalities (range: 0.60–0.65). Notably, extracardiac signals frequently appeared in renal regions, consistent with systemic amyloid burden reported in radiological literature (Kawashima et al., 2011).

4. Discussion and Conclusion

Our study demonstrates that a multimodal deep learning framework integrating cine, LGE, and mapping sequences can achieve strong performance for CA subtype classification. The xLSTM-based encoder strikes a favorable balance between accuracy and computational cost, making it suitable for potential clinical deployment.

Adding demographic information did not improve classification, potentially due to overfitting in limited sample sizes or the dominant discriminative power of imaging features. Furthermore, Grad-CAM activations in extracardiac regions may reflect systemic amyloid burden and open avenues for future research into holistic biomarkers.

These findings align with prior work emphasizing the value of CMR-derived tissue characterization (Fontana et al., 2015a; Banypersad et al., 2015). Our contributions extend this knowledge by providing an end-to-end interpretable framework that can learn from incomplete or heterogeneous imaging inputs.

In conclusion, our spatiotemporal model demonstrates strong performance and interpretability in CA subtype classification using multisequence MRI. Future work will focus on external validation with multi-institutional cohorts and deeper investigation into systemic imaging features.

References

Fabian Aus dem Siepen and Timon Hansen. Diagnosing AL and ATTR amyloid cardiomyopathy: A multidisciplinary approach. *Journal of Clinical Medicine*, 13(19):5873, October 2024. doi: 10.3390/jcm13195873.

Muhammad Awais Azam, Khalid Bashir Khan, Saim Salahuddin, Ehsan Rehman, Sheraz Ahmed Khan, Muhammad Adnan Khan, and Amir H. Gandomi. A review on multimodal medical image fusion: Compendious analysis of medical modalities, multimodal databases, fusion techniques and quality metrics. *Computers in Biology and Medicine*, 144:105253, 2022. doi: 10.1016/j.compbiomed.2022.105253.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014. doi: 10.48550/arXiv.1409.0473.

Sanjay M. Banypersad, Marianna Fontana, Viviana Maestrini, Daniel M. Sado, Gabriella Captur, Aviva Petrie, Stefan K. Piechnik, Carol J. Whelan, Anna S. Herrey, Julian D. Gillmore, Philip N. Hawkins, and James C. Moon. T1 mapping and survival in cardiac amyloidosis. *European Heart Journal*, 36(4):244–251, 2015. doi: 10.1093/eurheartj/ehu444.

Maximilian Beck, Korbinian Pöppel, Philipp Lippe, Ryna Kurle, Per M. Blies, Ludwig Ruff, Iman Salehi, Mohammad Afsari, Robert Slama, Keyan Brantley, Patrick van der Smagt, Jürgen Schmidhuber, and Sepp Hochreiter. *arXiv preprint arXiv:2405.04517*, 2024.

James Bergstra, Daniel Yamins, and David D. Cox. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. *Proceedings of the 30th International Conference on Machine Learning (ICML-13)*, 28:115–123, 2013.

Sharmila Dorbala, Sarah Cuddy, and Rodney H. Falk. How to image cardiac amyloidosis. *Circulation: Cardiovascular Imaging*, 13(2):e010435, 2020. doi: 10.1161/CIRCIMAGING.119.010435.

Marianna Fontana, Rick Chung, Philip N. Hawkins, and James C. Moon. Cardiovascular magnetic resonance for amyloidosis. *Heart Failure Reviews*, 20(2):133–144, 2015a. doi: 10.1007/s10741-014-9470-7.

Marianna Fontana, Silvia Pica, Patricia Reant, Amna Abdel-Gadir, Thomas A. Treibel, Sanjay M. Banypersad, Viviana Maestrini, Stefania Rosmini, Heerajnarain Bulluck, Ravin Sayed, Kalpa Patel, Saidi Mamhood, Chiara Bucciarelli-Ducci, Carol J. Whelan, Anna S. Herrey, Julian D. Gillmore, Philip N. Hawkins, and James C. Moon. Prognostic value of late gadolinium enhancement cardiovascular magnetic resonance in cardiac amyloidosis. *Circulation*, 132(16):1570–1579, 2015b. doi: 10.1161/CIRCULATIONAHA.115.016567.

Akira Kawashima, William G. Alleman, Naoki Takahashi, Bohyun Kim, Bernard F. King Jr, and Andrew J. LeRoy. Imaging evaluation of amyloidosis of the urinary tract and retroperitoneum. *RadioGraphics*, 31(6):1569–1582, 2011. doi: 10.1148/rg.316115510.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*, 2015. doi: 10.48550/arXiv.1412.6980.

Ze Liu, Jia Ning, Yue Cao, Yixuan Wei, Zheng Zhang, Stephen Lin, and Han Hu. Video swin transformer. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3202–3211, 2022. doi: 10.1109/CVPR52688.2022.00320.

Nicola Maggialetti, Natale Daniele Brunetti, Adriana Anna Stabile Ianora, Giuseppe Angelelli, Danilo Angiletta, Guglielmo Boccia, Vincenzo Triggiani, Pietro Scicchitano, Angela Piscitelli, Tiziano Valente, and Angela Sardaro. Role of cardiovascular magnetic resonance in cardiac amyloidosis: A narrative review. *Journal of Personalized Medicine*, 14(4):407, 2024. doi: 10.3390/jpm14040407.

Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-CAM: visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 618–626, 2017. doi: 10.1109/ICCV.2017.74.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, pages 5998–6008, 2017. doi: 10.48550/arXiv.1706.03762.

Lei Zhao, Zhuang Tian, and Quan Fang. Diagnostic accuracy of cardiovascular magnetic resonance for patients with suspected cardiac amyloidosis: a systematic review and meta-analysis. *BMC Cardiovascular Disorders*, 16:129, June 2016. doi: 10.1186/s12872-016-0311-6.