An Explainable Framework for Systolic Murmur Classification Using Graph Attention Network

Shib Sundar Banerjee[1], Francesco Renna[1]

[1] INESC TEC, Faculdade de Ciência da Universidade do Porto, Porto, Portugal

Abstract

Despite recent advances, deep learning for computer-aided diagnostics using heart sound is still limited in clinical transparency. To address that gap, this study presents an explainable Graph Attention Network (GAT) framework for automated cardiovascular disease diagnosis and systolic murmur characterization using phonocardiogram (PCG) signals. The system segments PCG signals into standard cardiac cycle events, encoding wavelet-derived time-frequency mappings into sequential heterographs via a Bi-GRU-based framework. A hierarchical GAT then classifies pathological systolic murmurs into multiple classes. Validated on the Yaseen dataset, the model achieved an accuracy of 100.00% *in murmur detection and* 98.50% *in multi-class classification. An integrated explainability module uses attention weights to provide pathological descriptions for detected murmurs. Combination of these explanations with classification confidence scores facilitates reliable clinical decision-making.*

1. Introduction

Phonocardiogram (PCG) signals have long served as the gold standard of clinical auscultation, acting as the first-line assessment for valvular heart diseases (VHDs) [1]. Signature acoustic anomalies, such as murmurs, are recognized as definitive diagnostic indicators due to their unique hemodynamic etiologies across VHDs, throughout cardiological literature. Precise identification and characterization of murmurs can yield vital clinical insights into the exact typology of the underlying disease, although it can be challenging even for human experts [2]. Building computational frameworks that can map and explain these acoustic anomalies is therefore key to establishing clinical transparency and diagnostic reliability.

Despite their clinical utility, traditional deep learning-based methods for automated PCG analysis present severe architectural and operational limitations. Most established pipelines process either raw 1D acoustic waveforms or their corresponding 2D time-frequency spectral representations, such as spectrograms or Mel-frequency cepstral coefficients (MFCCs) [3]. Although these models achieve high performance, they fail to provide a description or localization of the acoustic features driving the inference and scope for clinical validation. Furthermore, standard neural architectures suffer from a rigid input bottleneck.

To resolve these challenges, representing physiological signals as graph structures has emerged as a powerful paradigm. Graph-based approaches are uniquely suited for PCG analysis because they preserve the sequentiality of the acoustic events throughout the cardiac cycle while natively handling variable signal lengths [4]. Similar approaches have been successfully demonstrated in other complex cardiovascular domains, particularly in the structural representation of electrocardiograms (ECGs). Recent frameworks have shown that mapping ECG into spatio-temporal graphs for both uni-lead and multi-lead contexts enables models to achieve high diagnostic accuracy for diverse tasks while enabling better explainability [4, 5].

This study presents a Graph Attention Network (GAT) approach designed for an explainable clinical expert system, providing both automated diagnosis of cardiovascular disease and clinically relevant characterization of systolic murmurs. It involves time-frequency transformation of the signals, encoding the state-specific patches, developing a sequential heterograph representation, and finally utilizing a custom GAT Network for identification and diagnosis of systolic murmurs. The explainability outcomes are assessed to determine the alignment with the clinical labels for spectral biomarker discovery.

2. Methodology

2.1. Design of Graph Attention Network

The proposed framework integrates raw PCG signals with an automated segmentation step to achieve precise temporal localization of cardiac cycle events (S1, Systole, S2, Diastole). Wavelet-derived time-frequency mappings are encoded via an attention-enabled Bi-dimensional Gated Recurrent Unit to capture multiscale temporal dependencies and isolate prominent acoustic features. This block decouples scale and time by average-pooling the input across the temporal dimension to generate a scale-wise attention vector while average-pooling across the scale dimension to produce a step-wise temporal attention vector, which is subsequently used to construct a joint 2D importance map. Four different autoencoders are utilized for capturing the unique spectral representations of the four cardiac cycle events. Resolution is enhanced in systolic and diastolic regions as compared to the fundamental heart sounds (10 : 1), optimizing the characterization of spectral anomalies in targeted landmarks. The encoded vectors are subsequently transformed into sequential heterographs. In this scheme, each event is treated as a distinct node, rather than as a homogeneous graph, to preserve strict chronological order and enable state-specific attention. A hierarchical GAT system performs murmur detection followed by multi-class classification.

To model the complex temporal sequencing and physiological transitions of the cardiac cycle, a Heterogeneous Graph Attention Network (HeteroGAT) classifier is proposed that can handle the edge heterogeneity by preserving multi-relational configuration [6]. The network architecture consists of two stacked heterogeneous graph convolutional layers built upon the GATv2 formulation, which dynamically model directed edge transitions across a predefined set of cardiac cycle events. To preserve temporal sequence integrity, standard node self-loops are omitted. Instead, recursive transitions are modeled explicitly as directed homophilous edge types connecting consecutive instances of identical states across cycles. To transition from node-level representations to a graph-level signature, global mean pooling operator is employed across each active state node type, summing the resulting vectors to construct a unified global latent signal embedding.

Classification is executed via a dual-stage hierarchical design that mimics clinical reasoning. A primary linear detection head first provides a binary decision for murmur detection. These resulting logits are normalized via a Softmax function and concatenated directly with the original global signal embedding and ingested into a secondary linear disease head, computing confidence to classify the specific underlying disease. Finally, the framework maintains clinical interpretability by permitting the direct extraction of heterogeneous edge-specific attention weights from the first convolution layer, enabling the post hoc tracing of physiological state interactions that drive the diagnostic inference. A schematic of the pipeline with architectural information is provided in Figure 1.

2.2. Dataset and Training

The public database by Yaseen [7] is utilized for this pilot study due to the high-quality short-duration data it presents. Apart from the normal samples ($N = 200$), it includes other disease cases. The cases with the presence of recognized systolic murmurs are retained for this study: Aortic Stenosis (AS), Mitral Regurgitation (MR), and Mitral Valve Prolapse (MVP). There are again $N = 200$ samples for each of these disease classes. A train-test balanced split of $80 : 20$ has been prepared for model training and evaluation. It is to be noted that there is no ground truth segmentation of the cardiac cycle events for the given dataset. However, the sim2real approach has been utilized for obtaining precise segmentation of both normal and pathological signals, which have been presented in a prior work [8]. The signals alongside model-derived segmentation masks are used as the input for the GAT framework presented in the current study.

To optimize the proposed multi-task network, a joint optimization training strategy is employed to simultaneously learn binary murmur detection and multi-class classification. The model parameters are optimized end-to-end using the Adam optimizer with an initial learning rate of $\eta = 10^{-5}$ and a weight decay of 10^{-3} to provide L_2 regularization. Total loss function (L_{total}) is formulated as the unweighted sum of two independent cross-entropy objectives: $L_{\text{total}} = L_{\text{det}} + L_{\text{dis}}$, where L_{det} represents the cross-entropy loss for the binary detection head, and L_{dis} is the standard cross-entropy loss applied to the multi-class disease identification head. Optimization is initialized with a base learning rate of 1×10^{-5}, which is dynamically decayed by a factor of 0.5 whenever the validation loss plateaues for 20 consecutive epochs.

2.3. Attention Map-based Explainability

A post-hoc explainable framework is established that maps two-dimensional acoustic attention weights directly to time-frequency representations. For each identified segment, the pre-trained autoencoder extracts a temporal-spectral feature representation along with an associated 2D attention map ($\mathbf{W}_{\text{attn}} \in \mathbf{R}^{T\times F}$), where T and F denote the localized time steps and frequency scales, respectively. The framework implements a morphological profiling routine focused exclusively on the systolic phases. To decode the spectral-temporal representation, the 2D attention map of each systolic segment is collapsed along the frequency axis to yield a min-max normalized 1D temporal attention profile:$P_{\text{attn}}(t) = \text{Norm}\left(\frac{1}{F}\sum_{f=1}^{F}\mathbf{W}_{\text{attn}}(\mathbf{t},\mathbf{f})\right)$. From this 1D temporal profile, three shape-based features are extracted to characterize the murmur:

- **Peakiness:** The ratio of the maximum attention value to the mean attention value across the segment.
- **Spread:** The temporal duration over which the attention values exceed a prominent activation threshold.
- **Mean Time:** The centroid of the attention distribution across the normalized temporal duration.

A decision tree trained on the training data using these

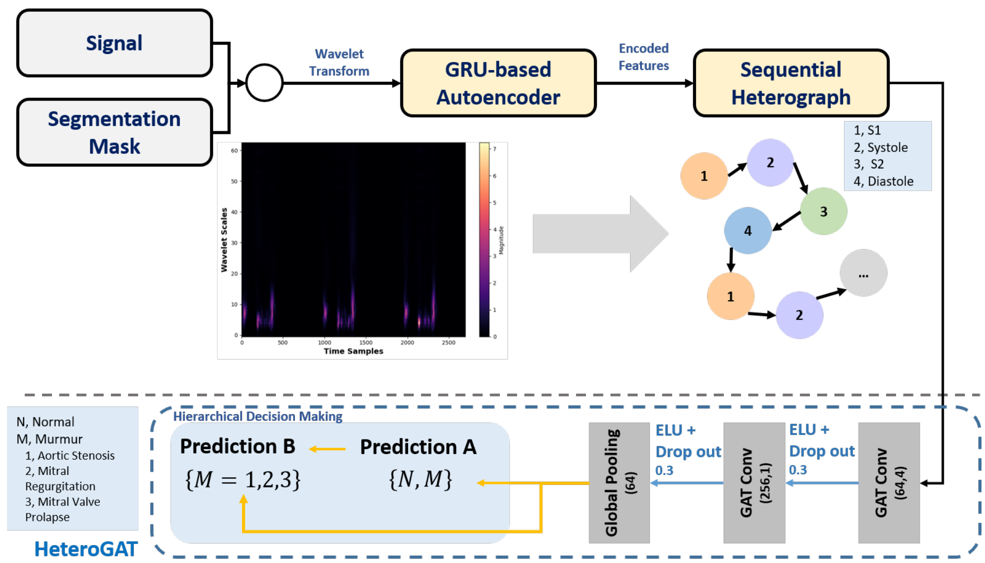

Figure 1. Schematic of the proposed pipeline for hierarchical classification of systolic murmur using the HeteroGAT framework. (GAT Conv: Attention-enabled graph convolutional block - version 2, ELU: Exponential linear unit)

shape-based features is further utilized for inference of the murmur characteristics on the unseen testing data, and the alignment with the clinical label is evaluated. The inference is made for each systolic interval, and a consensus is taken for overall decision for the murmur characterization.

3. Results and Discussion

3.1. Classification Outcomes

The model performance for the hierarchical decision making is demonstrated in the confusion matrix for the testing set, as shown in Figure 2A. The proposed model achieves 100% accuracy in terms of detection of the systolic murmurs. Further subclassification for disease cases also achieves a macro F1 score of 98%, which is comparable to the SOTA methods. Highest specificity is achieved for AS, followed by MR and MVP, with the respective values of F1 scores of 99.01%, 98.04%, and 96.91%. The lowest sensitivity is noted for MVP, but close observation reveals the variability of the acoustic manifestation of the usual click pattern, which can mimic other disease classes. The outcomes can further be verified using the explainability frameworks.

3.2. Explainability and Clinical Insights

The three chosen shape-based features show distinct ranges for the considered disease classes, and it is evident from the median fingerprint of each feature as depicted in Figure 2B. This demonstrates the possibility of utilizing the attention map-derived features for explainability framework. Figure 2C shows a narrative example of a sample signal for the model-derived decision and post-hoc interpretation using the attention map. The actual clinical label is MVP, and the model predicts the correct class with a confidence score of 98.99%. However, the inference derived from the attention map reveals that systolic murmur patterns vary considerably between cardiac cycles. While the first two murmurs present a weak click followed by a regurgitation pattern, the last murmur presents a strong click. The inference mechanism rightly identifies this variability by flagging "MR/MVP Mixed Pattern" for the first two murmur segments and "strong MVP click" for the last one. The consensus is the mixed pattern of MR/MVP.

4. Conclusion

This study presents an attention-enabled graph-based approach integrated with an interpretability framework for the detection and characterization of systolic murmurs. The model achieves 100.00% accuracy in the murmur detection task and macro accuracy of 98.50% and F1 score of 98.49% for the subclassification task. The inference for the spectral signature of murmurs can provide precise characteristics of each disease class and its respective manifestation, as specified in the narrative example. The future work would include extending this approach for a more complex, metadata-rich dataset like Physionet2016/CirCor for verifying the efficiency and discovery of specific spectral patterns for unique pathology cases.

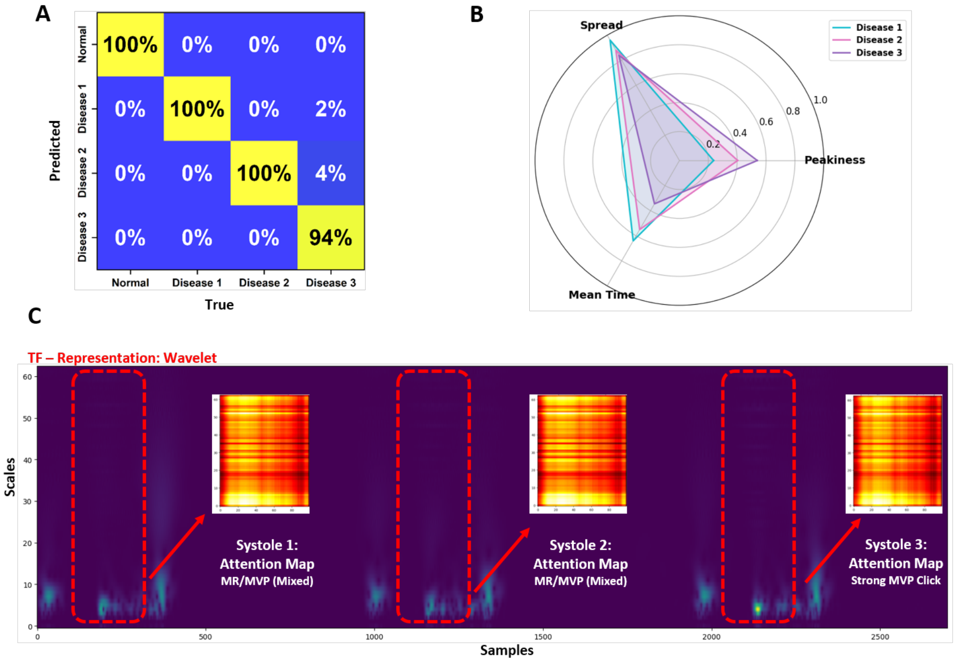

Figure 2. Classification and interpretability outcomes for the proposed HeteroGAT approach: A, Outcome for the multiclass classification (Disease 1: Aortic Stenosis, Disease 2: Mitral Regurgitation, Disease 3: Mitral Valve Prolapse), B, Normalized median of the shape-based features derived from attention maps for varied disease classes, C, Sample explainability for a representative case (Actual: Mitral Valve Prolapse, Predicted: Mitral Valve Prolapse, score: 98%) with overlying attention maps for systolic segments

Acknowledgments

This work is co-funded by the European Regional Development Fund (ERDF) through the Innovation and Digital Transition Programme (COMPETE 2030) under Portugal 2030 and by National Funds through the FCT - Fundação para a Ciência e a Tecnologia, I.P. (Portuguese Foundation for Science and Technology) within project PULSE, with reference 17172 (COMPETE2030-FEDER-00864900) (https://doi.org/10.54499/17172 (COMPETE2030-FEDER-00864900)).

References

[1] Olatunji DE, Zannu JD, Mukamakuza CP, Uiso GN, Buol C, Thuo JB, Ghogomu NT, Aman MMM, Umubyeyi E. Machine learning-based analysis of ECG and PCG signals for rheumatic heart disease detection: A scoping review (2015-2025). Comput Methods Programs Biomed Update 2025; 100228.

[2] Davidsen AH, Andersen S, Halvorsen PA, Schirmer H, Reierth E, Melbye H. Diagnostic accuracy of heart auscultation for detecting valve disease: a systematic review. BMJ Open 2023;13(3):e068121.

[3] Kannan A, Saikia MJm, Kumar S, Datta S. Detection of valvular heart diseases from PCG signals using machine and deep learning models–a review. IEEE Access 2025;.

[4] Lee M, Lim J, Kim J. ECG-GraphNet: Advanced arrhythmia classification based on graph convolutional networks. Heart Rhythm O2 2025;.

[5] Maurer MC, Hempel P, Steinhaus KE, Chereda H, Vollmer M, Krefting D, Spicher N, Hauschild AC. xGNN4MI: explainability of graph neural networks in 12-lead electrocardiography for cardiovascular disease classification. npj Digit Med 2026;.

[6] Lu K, Yu Y, Guo R, Cheng N, Huang Z, Ma Y, Liang M, Wang Y, Qin X, Ren Y, et al. Addressing graph heterogeneity and heterophily from a spectral perspective. Neurocomputing 2026;133500.

[7] Yaseen, Son GY, Kwon S. Classification of heart sound signal using multiple features. Appl Sci 2018;8(12):2344.

[8] Banerjee SS, Renna F. Evaluation of segmentation performance for pathological heart sounds using a sim2real deep learning approach. In Proc. 31st Portuguese Conf. Pattern Recognition. 2025; .

Address for correspondence:

Shib Sundar Banerjee
Faculty of Sciences, University of Porto
shib.banerjee@inesctec.pt