From Signals to Graphs: A Novel Approach for Intelligent ECG Audit

Guilherme H G Evangelista[1], Pedro B Rigueira[1], Victoria F Mello[1], Luisa G Porfírio[1], Artur X Nascimento[1], Caio S Grossi [1], Raquel X Teodoro[1], Gabriela M M Paixão[2] Gisele L Pappa[1], Antonio L Ribeiro[2], Wagner Meira Jr[1]

[1]Departamento de Ciência da Computação, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil
[2]Centro de Telessaúde, Hospital das Clínicas da UFMG, Belo Horizonte, Brazil

Abstract

Electrocardiograms (ECGs) are crucial for cardiac diagnostics, but current audit practices often employ random selection, resulting in inefficiency and overlooking valuable analytical insights. To optimize resource allocation and enhance audit outcomes, a targeted approach is needed, prioritizing ECGs with high diagnostic ambiguity or clinical relevance. This study introduces an automated method for identifying such high-utility ECGs using graph-based features derived from visibility graphs and time series features. By transforming ECGs into graph representations and using their time series properties, we enhance the discrimination of four key conditions: right bundle branch block (RBBB), left bundle branch block (LBBB), sinus tachycardia (ST), and sinus bradycardia (SB). These features capture complex, non-linear relationships in ECG data, enabling more nuanced clustering than conventional methods. Our approach not only streamlines medical audits by targeting the most informative ECGs but also demonstrates the untapped potential of graph theory combined with time series in cardiac diagnostics. By reducing reliance on random sampling, this method can improve audit efficiency by over four times, reducing costs, and ultimately supporting better clinical decision-making.

1. Introduction

The 12-lead electrocardiogram (ECG) is central to cardiac diagnosis, yet audit workflows often rely on random or convenience sampling, which yields few contentious cases and low diagnostic discordance (12%) [1]. Targeting exams that are more likely to be ambiguous can make audits markedly more informative. In prior work, we clustered ECG signal embeddings and increased the discordance rate to 38.9% [2]. However, latent representations offer limited clinical interpretability, limiting subsequent analysis and clinician confidence.

In this paper, we propose an alternative that combines graph and time-series features for targeted audit selection. We transform ECGs into visibility graphs (VGs) [3] and extract topological features, complemented by time and frequency statistics computed directly from the signals. While VGs have been explored mainly for supervised classification [4], their use to drive unsupervised selection of ambiguous exams for audits remains underexplored. Our central hypothesis is that this joint, feature space separates typical from atypical exam patterns better than dense embeddings alone, increasing the yield of diagnostically ambiguous cases and exposing an ambiguity gradient aligned with cluster cohesion. We validate this against a strong embedding baseline, showing improved identification of high-utility ECGs for review while preserving clear signal-to-feature mappings that support clinical interpretation.

2. Method

2.1. Dataset

We used the CODE-15% [5] dataset, a large public collection of 12-lead ECGs. Our analysis focused on four prevalent conditions: Sinus Tachycardia (ST), Sinus Bradycardia (SB), Right Bundle Branch Block (RBBB), and Left Bundle Branch Block (LBBB). We selected these four conditions because they are phenotypically coherent and amenable to unsupervised grouping, two sinus rhythm variants and two conduction block patterns. For each condition, we created a balanced dataset containing all available exams with that single label, plus a set of 50,000 ECGs labeled as normal sinus rhythm, following the methodology of our previous work.

2.2. Feature Engineering

To create a comprehensive representation of each ECG, we engineered a hybrid feature set designed to capture complementary aspects of cardiac dynamics. This set com-

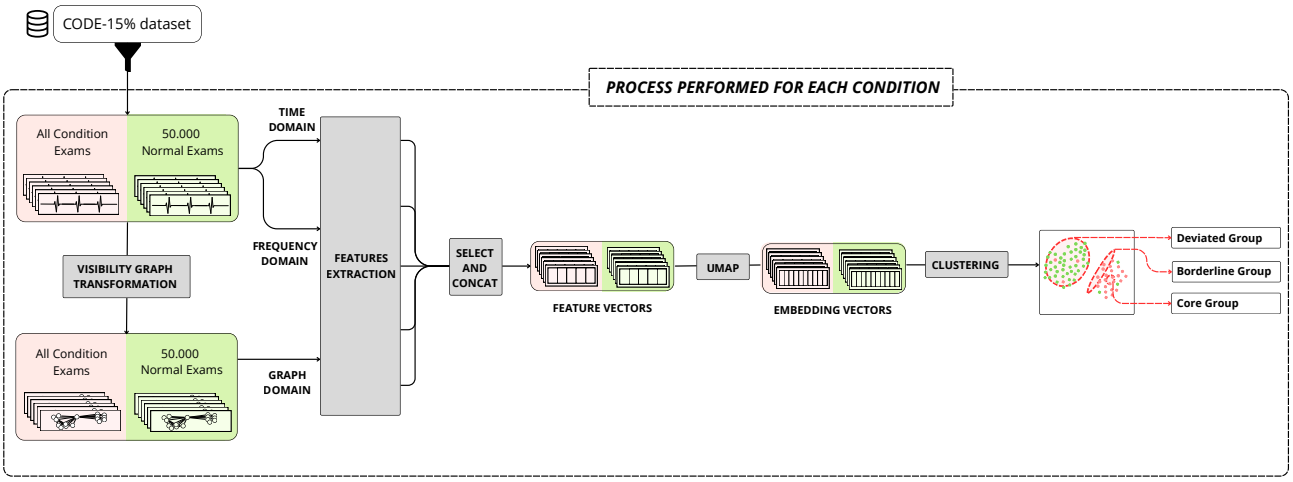

Figure 1. End-to-end pipeline for targeted ECG audit selection.

bines topological properties derived from visibility graphs with classical statistical descriptors from the raw time series. For each exam, features are extracted on a per-lead basis and later concatenated to form the final high-dimensional representation.

2.2.1. Graph-Based Topological and Structural Descriptors

Our primary innovation lies in transforming each ECG lead's time series into a visibility graph (VG), a method that maps temporal patterns into a network topology [3]. Each sample point (t_i, y_i) in the signal becomes a node, and an edge connects two nodes (t_a, y_a) and (t_b, y_b) if they have direct "line-of-sight," as defined by:

$$y_c < y_b + (y_a - y_b)\frac{t_b - t_c}{t_b - t_a}, \quad \forall c \text{ s.t. } t_a < t_c < t_b$$

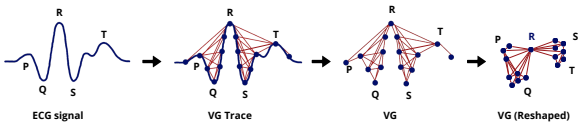

Figure 2. Illustration of the Visibility Graph transformation.

From each generated VG, we compute a rich suite of descriptors. These metrics are chosen to quantify the complexity, connectivity, and regularity of the underlying ECG signal. They are grouped into three main categories:

- **Unweighted Topological Metrics:** Statistics describing the graph's fundamental structure, including the mean, standard deviation, skewness, and kurtosis of the node degree distribution, graph density, and assortativity. These features capture the overall connectivity patterns.
- **Weighted Graph Metrics:** Descriptors that incorporate the Euclidean distance between connected nodes as edge weights. This category includes statistics on edge weight and node strength distributions, providing insight into the magnitude of temporal and voltage variations.
- **Regularity and Community Structure Proxies:** Metrics such as the global clustering coefficient and degree entropy, which serve as proxies for signal regularity and the presence of repeating, structured patterns within the ECG waveform.

2.2.2. Statistical Time and Frequency-Domain Descriptors

To complement the topological information, we extract a set of well-established statistical features directly from the preprocessed time series of each lead. These descriptors provide a foundational view of the signal's morphology and rhythm.

In the time domain, we characterize the signal's amplitude distribution and morphology using the mean, standard deviation, skewness, kurtosis, and range (min/max) of the signal amplitude. We also compute the mean and standard deviation of the first derivative to capture the rate of change. Rhythm regularity is assessed via the coefficient of variation of RR-intervals.

In the frequency domain, we analyze the spectral power distribution derived from Welch's periodogram. Key features include the power ratios in the low-band (0–15 Hz) and mid-band (15–50 Hz), the spectral centroid, and the spectral entropy, which together describe the concentration and complexity of the signal's frequency components.

The complete feature vector for an exam is formed by concatenating these descriptors across all 12 leads, resulting in a final representation with 432 features (36 fea-

tures/lead $\times$ 12 leads).

2.3. Feature Selection with AUC

The high-dimensional feature space, resulting from the combination of graph-based (228) and time-series (204) descriptors, required dimensionality reduction to avoid noise and redundancy. We adopted a univariate strategy based on the Area Under the ROC Curve (AUC), computed separately for each pathology (ST, SB, RBBB, LBBB). Features were ranked by their Separation Power, $|\mathrm{AUC} - 0.5|$, which reflects how well a variable distinguishes pathological from normal exams, independent of whether higher or lower values indicate disease. This ensured that both positively and negatively associated features were valued equally.

To further refine the set, we removed redundant variables by iteratively discarding those with strong correlation ($|r| > 0.90$) to higher-ranked ones, ensuring diversity and reducing multicollinearity. From the resulting ranked lists, the six most informative and non-redundant features were retained for each pathology. Finally, all selected features were standardized using z-scoring, so that they contributed equally to subsequent distance-based clustering.

2.4. Dimensionality Reduction and Clustering

The selected feature vectors first underwent dimensionality reduction using Uniform Manifold Approximation and Projection (UMAP)[6]. Subsequently, the low-dimensional embeddings were clustered using HDBSCAN[7], a density-based algorithm adept at identifying clusters of varying shapes and sizes while effectively handling noise. For each pathology, its corresponding feature vectors were clustered alongside a baseline set of 50,000 vectors from normal exams.

Our central hypothesis is that exams belonging to a specific pathology, but assigned by HDBSCAN to the predominantly "normal" cluster, represent atypical or clinically ambiguous cases. To ensure optimal performance, the hyperparameters for both UMAP and HDBSCAN were systematically tuned using the Optuna framework to maximize key clustering validation metrics.

To quantitatively evaluate the performance of our clustering model, we used the F1-score as an external validation metric, assessing the alignment between the generated clusters and the ground-truth diagnostic labels in the database. Defined as the harmonic mean of precision and recall, the F1-score provides a balanced measure of the algorithm's ability to correctly assign exams to their respective pathological or normal classes.

2.5. Clinical Validation of the Ambiguity Gradient

Beyond the quantitative metrics, our main objective was to assess the clinical relevance of the cluster structure. We hypothesized that diagnostic ambiguity is inversely related to an exam's cohesion within its pathological cluster. To test this, we adopted a three-tiered sampling strategy based on the membership probabilities assigned by HDBSCAN.

We selected 351 diseased ECGs across four pathologies (84 LBBB, 81 RBBB, 88 SB, and 98 ST), stratified into three groups:

1. **Core Group (n = 165):** High-probability cases within the pathological cluster (36 LBBB, 39 RBBB, 25 SB, 65 ST).
2. **Borderline Group (n = 100):** Exams at the cluster fringe, closer to the normal group (40 LBBB, 20 RBBB, 20 SB, 20 ST).
3. **Deviated Group (n = 86):** Cases misclassified into the normal cluster, representing the most ambiguous exams (8 LBBB, 22 RBBB, 43 SB, 13 ST).

Each group was independently re-evaluated by senior cardiology students, and the clinical discordance rate was computed as the percentage of diagnoses diverging from the original database labels. This stratified design enabled a direct and interpretable assessment of the proposed ambiguity gradient across the cluster structure.

3. Results and Discussion

We first evaluated our model's ability to separate pathological exams from normal ones. As shown in Table 1, the graph and time-series features significantly outperform the embedding-based baseline for conditions characterized by distinctive morphological changes. For RBBB and LBBB, the F1-Scores increased to 0.9639 and 0.9691, respectively. This indicates that topological descriptors from visibility graphs capture the structural signal distortions typical of conduction blocks more effectively than latent embeddings.

In contrast, for rhythm-based conditions such as Sinus Bradycardia (SB) and Sinus Tachycardia (ST), our method reached F1-Scores comparable to or slightly below the baseline. This suggests that embeddings may be more effective at capturing differences driven by a single global attribute like heart rate, whereas our broader feature set reflects finer variability within each group. Although this leads to lower separation in clustering, it also highlights the presence of ambiguous or borderline cases, precisely the type of exams that benefit most from targeted audit, as examined in the next analysis.

Table 1. Clustering F1-Score comparison.

Condition	Baseline	Selected Features
ST	0.9338	0.9180
SB	0.9146	0.7793
RBBB	0.9173	0.9639
LBBB	0.9367	0.9691

These borderline cases were then assessed in a dedicated clinical validation, summarized in Table 2. The results confirm a distinct ambiguity gradient: the Core group, representing typical cases, demonstrated high clinical concordance rates across all conditions (75.4% to 87.2%). As expected, the Borderline group showed intermediate and varied concordance (35.0% to 55.0%), confirming its status as a collection of structurally ambiguous cases.

The most significant result lies in the analysis of the Deviated group. The concordance for this group was exceptionally high for Sinus Bradycardia (SB) at 69.8% and for Sinus Tachycardia (ST) at 53.8%. This directly correlates with their lower F1-scores and demonstrates that our method successfully flags a concentrated subset of diagnostically challenging cases that are ideal for expert review.

Table 2. Clinical Concordance Rates (%) Across Stratified Ambiguity Groups, based on the clinical audit.

Condition	Core Group (Concordance)	Borderline (Concordance)	Deviated Group (Concordance)
LBBB	80.6%	35.0%	12.5%
RBBB	87.2%	55.0%	31.8%
SB	80.0%	50.0%	69.8%
ST	75.4%	40.0%	53.8%

To contextualize these findings, we directly compare our results against the embedding-based baseline from our previous work, which achieved an average discordance rate of 45.78% for the four abnormalities. Our proposed method demonstrates a marked improvement. The Borderline group, which represents exams at the fringe of pathological clusters, yielded an average discordance rate of 55.0%. More strikingly, the Deviated group achieved an even higher average discordance rate of 58.03%. This outcome is crucial as it quantitatively validates that our graph and time-series feature set surpasses the performance of dense embeddings in identifying diagnostically contentious cases. Furthermore, it confirms the existence of a practical ambiguity gradient. By targeting the Deviated group, auditors can focus on a highly concentrated pool of exams where diagnostic disagreement is highest, maximizing the efficiency and learning potential of the review process.

4. Conclusion

This paper introduced a novel method for intelligent ECG audit selection by combining visibility graph topology with time-series descriptors. Our approach successfully creates an"ambiguity gradient," identifying a concentrated subset of exams with high diagnostic uncertainty. The exams most likely to be misclassified by our model (the Deviated group) yielded a clinical discordance rate of 58.03% upon expert review. This result is a marked improvement over the 12% discordance typical of traditional random sampling, making our targeted selection process more than four times more efficient at surfacing the most clinically relevant cases for discussion. By focusing expert attention where it is most needed, our framework offers a simple, effective, and clinically justifiable route to more efficient and impactful ECG auditing. Future work will evaluate approximate graph constructions and alternative density estimators and extend to additional diagnoses.

Acknowledgments

This work is partially supported by CNPq, CAPES, Fapemig, as well as projects CIIA-Saúde and IAIA - INCT on AI.

References

[1] Gomes PR, et al. Sistema de laudos de eletrocardiograma: a importância de ferramentas de suporte à decisão 2020;.

[2] Rigueira P, Evangelista G, Porfírio L, Grossi C, Buzelin A, Pappa G, Paixão G, Ribeiro A, Jr. WM. Optimizing ecg audits: Clustering-based identification of ambiguous exams. In Anais do XXI Encontro Nacional de Inteligência Artificial e Computacional. Porto Alegre, RS, Brasil: SBC. ISSN 2763-9061, 2024; 61–72.

[3] Lacasa L, Luque B, Ballesteros F, Luque J, Nuño JC. From time series to complex networks: The visibility graph. Proceedings of the National Academy of Sciences April 2008; 105(13):4972–4975. ISSN 1091-6490.

[4] Kutluana G, Türker Classification of cardiac disorders using weighted visibility graph features from ecg signals. Biomedical Signal Processing and Control 2024;87:105420. ISSN 1746-8094.

[5] Ribeiro ALP, Paixão GM, Gomes PR, Ribeiro MH, Ribeiro AH, Canazart JA, Oliveira DM, Ferreira MP, Lima EM, de Moraes JL, Castro N, Ribeiro LB, Macfarlane PW. Tele-electrocardiography and bigdata: The code (clinical outcomes in digital electrocardiography) study. Journal of Electrocardiology 2019;57:S75–S78. ISSN 0022-0736.

[6] Leland M, John H, James M. Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv180203426 2018;.

[7] McInnes L, Healy J, Astels S, et al. hdbscan: Hierarchical density based clustering. J Open Source Softw 2017; 2(11):205.

Address for correspondence:

Guilherme H G Evangelista

Universidade Federal de Minas Gerais

guilherme.evangelista@dcc.ufmg.br