Optimizing ECG Audits: Clustering-Based Identification of Ambiguous Exams

Pedro B. Rigueira*[1], Guilherme H. G. Evangelista*[1], Luisa G. Porfírio*[1], Caio Souza Grossi*[1], Arthur Buzelin[1], Gisele L. Pappa[1], Gabriela M. M. Paixão[2], Antonio Ribeiro[2], Wagner Meira Jr.[1]

[1] Departamento de Ciência da Computação – Universidade Federal de Minas Gerais
[2] Centro de Telessaúde – Hospital das Clínicas da UFMG

Belo Horizonte – MG – Brazil

{pedrobacelar.rigueira, guilherme.evangelista}@dcc.ufmg.br

{luisagontijo, caio.grossi, arthurbuzelin, glpappa, meira}@dcc.ufmg.br

gabimiana@gmail.com, tom@hc.ufmg.br

Abstract. *Electrocardiograms (ECG's) are crucial tools for diagnosing heart diseases, and regular audits of these exams are essential to maintain diagnostic consistency, ensure quality standards, and secure the reliability of medical databases. However, the current practice of randomly selecting ECG's for audit can be inefficient, as it often includes cases with clear, uncontroversial diagnoses. In this paper, we present an unsupervised method that uses clustering techniques to identify ECG's with a higher likelihood of diagnostic ambiguity. Our approach identifies a group of exams with an average ambiguity rate of 38,98%, which is over three times higher than the 12% observed in conventional audit methods.*

1. Introduction

The 12-lead electrocardiogram (ECG) is an essential tool in modern medicine for detecting heart abnormalities. It records the heart's electrical activity, generating wave traces characterized by duration, amplitude, and morphology. The exam involves the strategic placement of 12 electrodes on the patient's body, each capturing electrical activity from a specific angle of the heart. The composite of signals captured from these directions, known as leads, provides a detailed view of the heart's electrical processes, making the ECG indispensable in both preventive and diagnostic medicine. To maintain diagnostic reliability and consistency, periodic medical audits are conducted [Johnston et al. 2000], during which a team of doctors discuss selected ECG's and review prior diagnoses. These audits also enhance the quality of medical databases by ensuring their precision and credibility — factors that are crucial for cardiology studies due to the complexities involved in cardiovascular diseases.

However, the effectiveness of current audit practices is often reduced by the reliance on random and convenience sampling to select ECG's [HSE 2013]. Random sampling ensures each unit has an equal chance of being chosen, while convenience sampling

*These authors contributed equally to this work.

selects based on ease of access [Bujang et al. 2012]. Therefore, these current processes are likely to include large amounts of ECG's with uncontroversial, widely agreed-upon diagnoses, which, while still useful, do not directly address the primary goal of an audit: identifying and debating potential diagnostic errors and ambiguities. Consequently, the efficiency of these audits may be compromised, as ECGs that can lead to more useful discussions and opportunities for improvement may not be reviewed. When using samples from these traditional techniques, audits show an average discordance rate of 12% among medical reports [Gomes et al. 2020], suggesting that only a small proportion of ECGs examined have real potential to provide valuable insights.

In regards to the current sampling methods and its limitations, we propose an innovative method to optimize the selection process of ECGs to be audited. Following the assumption that ECGs associated with a given disease are similar to each other, we cluster the exams based on their signals similarities. Evidently, the cluster containing the most ECG's for a given disease will include signals that exhibit the characteristic patterns of that illness, while those that fall outside this cluster are considered deviations from the typical pattern. Our solution utilizes this clustering strategy to identify the most suitable candidates for audit, specifically targeting on the deviated diseased exams and interpreting these misclusterings as a higher likelihood of ambiguities.

In this work, we generated audit samples for sinus tachycardia (ST), sinus bradycardia (SB), atrial fibrillation (AF), right bundle branch block (RBBB), and left bundle branch block (LBBB). Five cluster analyses were conducted to find abnormal ECGs that diverged from the typical disease pattern, using preprocessing steps like filtering and dimensionality reduction with autoencoders and UMAP [Leland et al. 2018], as further detailed in the Methodology section.

Our results indicate that, for ECG samples showing deviations from the pattern cluster, the discordance between the diagnoses recorded in the database and those made by senior cardiology students increases significantly. With an average of 38,98% of discordance, this number is over 3 times higher than the 12% discordance observed in regular audits. Such a result supports our hypothesis that these samples are particularly ambiguous or more challenging to diagnose and are worth auditing to understand the diagnostic discrepancies.

2. Related work

In this section, we present previous work that explored dimensionality reduction and machine learning techniques related to ECG analysis.

2.1. Machine Learning in ECG Analyses

Most existing studies on machine learning methods for ECG analysis tend to focus on highly specific applications and heavily rely on feature extraction, which leads to a loss of generality on the cardiac conditions analysed [Roopa and Harish 2017]. These features and ECG Patterns [Banerjee and Mitra 2014] are also largely studied, and play a specially significant role in supervised learning algorithms.

In terms of learning methods, supervised learning has been used the most for ECG analysis [Lyon et al. 2018], where models are trained on labeled data to detect and classify specific cardiac anomalies. Some of those focused on an improved generalization

capability [Ribeiro et al. 2020, de Chazal et al. 2004], or real-time patient-specific classification [Kiranyaz et al. 2016]. On the other hand, unsupervised learning is a more recent and evolving method that doesn't rely on labeled data, but, instead, aims to uncover hidden patterns or structures within the ECG data. For instance, ECG clustering aims to group signals based on similar characteristics, potentially revealing unexplored connections between various cardiovascular conditions [Nezamabadi et al. 2023].

2.2. Dimensionality Reduction

Dimensionality reduction plays a crucial role when working with large-scale datasets, often serving to preserve the most important features while reducing the complexity of the data. Traditional techniques, such as Singular Value Decomposition (SVD) and Principal Component Analysis (PCA), remain widely used in academic research due to their simplicity and effectiveness.

However, when dealing with highly complex datasets, these linear methods may fall short in capturing the essential characteristics of the data. In such cases, more advanced approaches have emerged to overcome these limitations, including neural networks, specifically autoencoders [Fournier and Aloise 2019], and more sophisticated algorithms like UMAP [Leland et al. 2018], which can offer superior performance in preserving critical data features.

2.3. Research Gap

Traditional sampling methods, such as random and convenience sampling, have been effective in specific contexts. However, these techniques tend to select consensual and less challenging ECGs, which may limit the potential of audits to identify and discuss more complex and error-prone cases. Although some recent studies have explored clustering techniques to identify patterns in ECGs, there is a significant gap in applying these techniques to optimize sample selection for medical audits. This work proposes to mitigates this problem by introducing a new methodology for sample selection in ECG audits, employing clustering to identify exams that deviate from expected patterns. This approach aims not only to enhance the efficiency of audits, but also to provide valuable insights into the causes of diagnostic discrepancies, contributing to improved diagnostic practices and the quality of medical databases.

3. Dataset

This work utilizes a subset of the CODE (Clinical Outcomes in Digital Electrocardiology) database [Ribeiro et al. 2019]. The CODE database is a 12-lead ECG dataset that includes anonymized patient information, clinical history, electrocardiographic diagnostic, among other data. It was a work developed with the database of digital ECG exams of the TeleHealth Network of Minas Gerais (TNMG) that brings out data collected between 2010 and 2016 of 2,325,114 exams from 1,558,415 patients. Each exam contains 12 leads with the labels DI, DII, DIII, AVR, AVL, AVF and V1-6. These labels carry signals representations of electrical activity from specific angles of the heart.

For the CODE, a hierarchical free-text machine learning algorithm was used to identify specific ECG diagnoses from cardiologist reports. Also automatic reports were generated using the Glasgow Diagnostic Statements [Macfarlane et al. 1990] or the Minnesota Code [Macfarlane and Latif 1996]. An exam was labeled with an abnormality in

the database when there was agreement between the cardiologist's extracted report and one of the automatic diagnoses. In cases where discrepancies occurred between the medical report and one of the automatic programs, a manual review was conducted by trained staff.

The CODE-15% subset [Ribeiro et al. 2021] [Lima et al. 2021] includes 345,779 exams from 233,770 patients, representing 15% of the CODE dataset. The exams were stratified by age, ranging from 16 to 85 years. An equal number of samples from each age group were randomly selected, ensuring a nearly uniform age distribution across the dataset.

The dataset includes six types of diseases, which were considered to representative of both rhythmic and morphological ECG abnormalities. This paper focuses primarily on five of these abnormalities: right bundle branch block (RBBB), left bundle branch block (LBBB), atrial fibrillation (AF), sinus tachycardia (ST), and sinus bradycardia (SB). The clustering process was particularly challenging for 1st degree AV block (1dAVb), so this condition will be set aside for future research.

3.1. Data Selection

With the CODE-15% dataset, we began by removing ECG's with interference. This was achieved by checking the medical reports and excluding any exams that had "interference" mentioned in their diagnoses. Following this, we randomly isolated 20,000 ECG's that did not show any of the six abnormalities. This latter set of ECGs was used in all five clusterings performed. We then eliminated all signals with multiple labels or no labels at all. For each clustering process, we utilized only the ECG's that had a definitive, single label. This process was detailed in Figure 1.

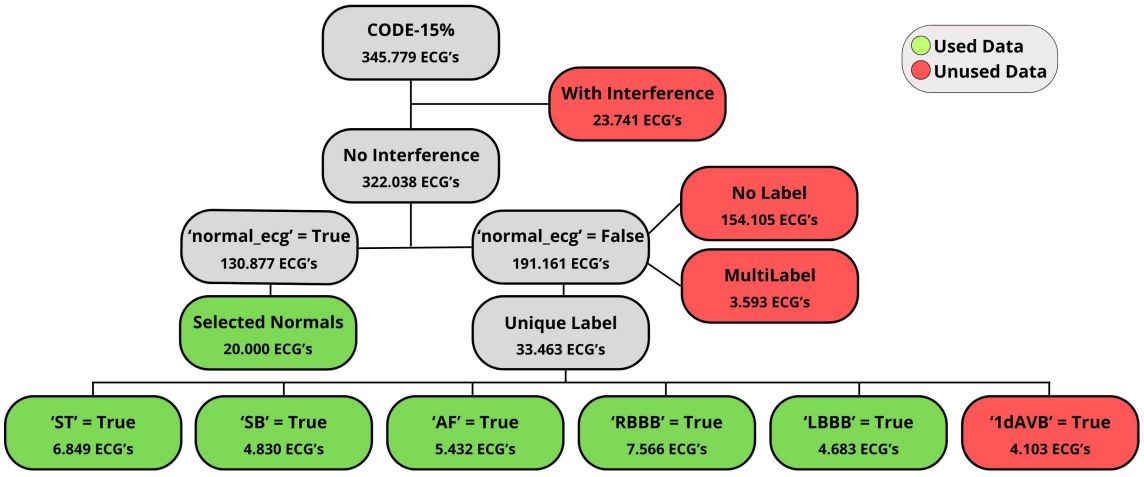

Figure 1. Data selection process.

4. Methodology

In this section, we will explore the methods employed to obtain our results, emphasizing their key characteristics and explaining their significance in the process.

4.1. Pre-Processing

In our work, a significant challenge lies in the extraction of features that most accurately represent ECG signals for clustering purposes. Consequently, it was essential to implement a robust filtering process to ensure that the algorithm did not interpret noisy information beyond the intended scope.

To achieve this, a high-pass filter [Berkaya et al. 2018] with a cutoff frequency of 0.5 Hz was employed. This filter is specifically designed to eliminate low-frequency components, such as baseline wander, which can be introduced by patient movement or respiratory activities, leading to a gradual drift in the ECG signal's baseline. By allowing only frequencies above 0.5 Hz to pass, the filter effectively preserves critical cardiac signal features while removing noise that may distort the analysis. This filtering process was systematically applied to all leads of each ECG signal before any further processing or input into neural networks. By doing so, the preprocessing ensured that only the most relevant and clean information was retained for subsequent stages of the analysis.

4.2. Dimensionality Reduction

Dimensionality reduction was a critical aspect of our work, enabling us to effectively manage and analyze high-dimensional data. We employed two methods for this purpose: an autoencoder and UMAP, as represented in Figure 2.

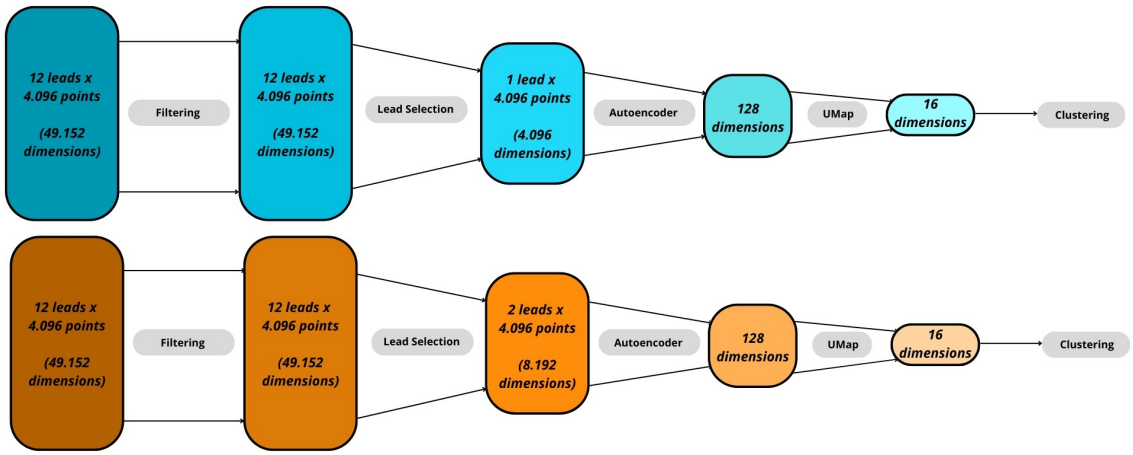

Figure 2. Dimensionality reduction process.

4.2.1. Autoencoder

We began by using an autoencoder, a powerful tool known for its ability to simplify complex data into a more manageable form with minimal loss of important details [Wang et al. 2014]. This network operates with an encoder that reduces the dimensionality of the original data and, with a decoder, attempts to reconstruct it as accurately as possible. This approach allowed us to handle the complexities of reducing dimensions while keeping the data's integrity as faithful as possible for further analysis.

In our work, we developed three autoencoders to capture the fundamental characteristics for each cardiac condition. We trained these autoencoders using a selective

set of leads from the ECG, aiming not only to emphasize the most informative leads but also to serve as a complementary dimensionality reduction technique (Figure 2). Initially, the approach was designed to replicate a medical professional's focus on the most diagnostically relevant leads. We further optimized the autoencoders by testing various lead combinations to enhance both reconstruction and clustering results. The most effective combinations found were: the D2 lead alone for atrial fibrillation (AF), sinus tachycardia (ST), and sinus bradycardia (SB); and leads V1 and V6 for right bundle branch block (RBBB) and leads V1 and V4 for left bundle branch block (LBBB). Notably, the effectiveness of lead V4 for LBBB was unexpected, as it is not commonly used for diagnosis.

The autoencoders were designed using a Fully Convolutional Network (FCN) [Chiang et al. 2019]. An FCN is a specialized version of traditional CNNs where dense layers are replaced with convolutional layers. This design choice helps preserve the locally-spatial information of neighboring input regions, which fully connected layers typically struggle to maintain. Additionally, by avoiding pooling layers, the network retains detailed structural and textural information that might otherwise be lost. We employed exponential linear units (ELU) as activation functions for the hidden layers, while the output layer in the FCN model has no activation function. Furthermore, each hidden layer is equipped with batch normalization to improve training stability and performance.

To ensure the accuracy of this process, we used the Mean Absolute Error (MAE) as our loss function. MAE calculates the average of the absolute differences between the original and reconstructed data points, providing a straightforward measure of reconstruction accuracy.

For validation purposes, we used the root mean square error (RMSE) as the evaluation metric. RMSE calculates the variance between the original and reconstructed values across all points in the exam, providing a quantitative measure of the reconstruction accuracy, with lower variance indicating better reconstruction. The RMSE presented in Table 1 was calculated by taking the mean of the RMSE values from all reconstructions for each exam. In the RMSE formula (Equation 1), N represents the number of data points in a specific exam. The variable x_i denotes the value of point i in the original signal, while $\hat{x}_i$ represents the value of point i in the exam reconstructed by the model.

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2} \tag{1}$$

Table 1. Autoencoders RMSE.

Autoencoder Type	One Lead D2	Two Leads V1 & V6	Two Leads V1 & V4
RMSE	0.0289	0.0512	0.0420

4.2.2. UMAP

We use UMAP (Uniform Manifold Approximation and Projection) [Leland et al. 2018] as a complementary technique to the autoencoder. UMAP is distinguished by its ability to

preserve both the local and global structure of the data, which is crucial for complex and non-linear datasets, such as ECG signals. After the initial reduction to 128 dimensions by the autoencoder, we apply UMAP to further reduce the dimensionality to 16, due to curse of dimensionality [Assent 2012]. This step is vital for optimizing the clustering process, enabling more effective analysis with minimal information loss.

UMAP begins its reduction by calculating the distances and connectivity probabilities between points in the original space, using a logarithmic function to determine local density. It then performs an iterative optimization to project the data into a lower-dimensional space, maximizing the similarity between nearby points and minimizing the connectivity between distant points through a cost function based on Kullback-Leibler (KL) divergence.

The essential parameters that influence effective dimensionality reduction with UMAP are the number of neighbors and the minimum distance between points. The number of neighbors adjusts the balance between preserving local and global structure: low values highlight local structure, while high values provide a broader view of the data. The minimum distance defines the minimum separation between points in the reduced projection, affecting the density of clusters: low values create more compact clusters, while high values preserve the overall topological structure. We empirically adjust these parameters, testing different configurations and visualizing the projections to determine the best conditions for clustering.

To validate the effectiveness of the reduction, we used the trustworthiness metric [Pedregosa et al. 2011], which assesses how well the local structure of the original data is preserved in the low-dimensional projection. This metric ranges from 0 to 1, with values close to 1 indicating higher preservation of neighborhood relationships.

Trustworthiness is calculated by considering the total number of samples n and the k nearest neighbors of each sample i. The ranking $r(i, j)$ of each neighbor j of i in the original space is compared with its position in the low-dimensional projection. If the position of j changes significantly in the projection, the difference is penalized, reducing the trustworthiness value.

$$T(k) = 1 - \frac{2}{nk(2n - 3k - 1)} \sum_{i=1}^{n} \sum_{j \in {}_{k}^{N_i}} \max\left(0, (r(i, j) - k)\right) \tag{2}$$

Equation 2 presents the formula used for this calculation, and the results, shown in Table 2, indicate that the proposed reduction maintains high trustworthiness, ensuring that the essential structure of the data has been preserved.

Table 2. UMAP Trustworthiness.

UMAP	One Lead D2	Two Leads V1 & V6	Two Leads V1 & V4
Trustworthiness	0.9865	0.9733	0.9419

4.3. Clustering

With the data now reduced to 16 dimensions, we can proceed to the clustering phase. While several algorithms may be employed at this stage, Hierarchical Density-Based Spa-

tial Clustering of Applications with Noise (HDBSCAN) [McInnes et al. 2017] was chosen due to its suitability for our type of data and problem. HDBSCAN excels in handling noise and outliers, does not require prior knowledge of the number of clusters, and can accommodate clusters of varying shapes. This makes it particularly well-suited for ECG signals, which often contain noise and have uncertain features. The considered distance between each tracing was the standard euclidian distance.

For each condition evaluated, a separate experiment was conducted. We selected all ECG's associated with the condition and combined them with 20000 normal ECG's as inputs for the algorithm. Running HDBSCAN on this mixed data consistently resulted in the formation of two dense clusters. Details of the specific data combinations used for each of the five clustering experiments are provided in Figure 3.

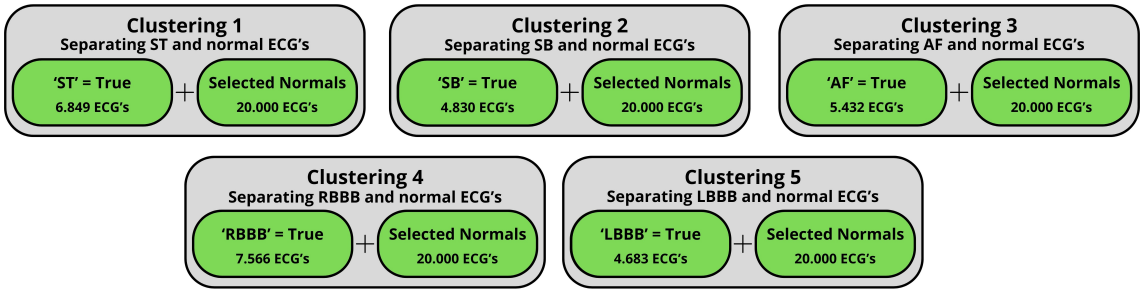

Figure 3. Data selection for each clustering perfomed.

The HDBSCAN parameters were empirically tuned to prevent the formation of overly specific clusters or the collapse into a single cluster. This careful adjustment allowed us to maintain two dense, well-separated clusters. The clear separation between these clusters suggests that, in most cases evaluated, diseased ECG's can be distinctly differentiated from normal ones using this representation. In all experiments, one group is consistently more dispersed, representing the predominantly normal group, while the other group is quite centralized, representing the ECG's with the evaluated conditions. This reflects the diverse and less-defined nature of normal electrocardiograms compared to the more specific patterns of diseased signals.

4.4. Medical Analysis

To further investigate the misclustered signals, we compared them with ECG's that were accurately clustered and asked cardiologists to classify these two groups. Our goal was to determine whether the agreement between specialists and the CODE-15% classification differed between correctly and incorrectly clustered diseased ECG's. We provided cardiologists with 45 samples of correctly clustered diseased ECG's from each cluster, as well as misclassified samples: 37 ST's, 12 SB's, 17 AF's, 22 RBBB's, and 22 LBBB's.

For each evaluated ECG, cardiologists had the following classification options: one of the six diseases listed in CODE-15%, another disease, no disease, or unable to evaluate due to interference. None of the samples included the CODE-15% labels to avoid introducing bias into the analysis and compromising the integrity of the results.

5. Results and Discussion

In the analyses of the five diseases, the clustering process generated two distinct clusters for each: one representing ECG's with the specific disease and another comprising ECG's

considered not to have that disease. The visual representation of the clustering results using the first two dimensions of UMAP is shown in Figure 4. The characteristics of each cluster is detailed in Table 3.

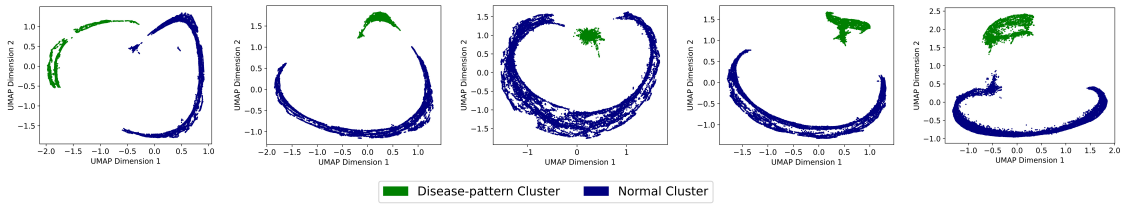

Figure 4. Clustering results: ST, SB, AF, RBBB and LBBB, respectively.

Table 3. Distribution of normal and diseased ECG's in Clusters.

Clustering	**ST**		**SB**		**AF**		**RBBB**		**LBBB**	
Label	Normal	ST	Normal	SB	Normal	AF	Normal	RBBB	Normal	LBBB
Cluster A	231	6201	765	4715	447	3614	898	7171	126	4237
Cluster B	19769	648	19235	115	19553	1818	19102	395	19874	446

Using the labels we had before clustering, we externally validated the purity of these clusters, which is reflected in the F1 score (see Table 4)[Pedregosa et al. 2011]. The F1 score provided an indication of how well the clustering algorithm separated diseased from non-diseased ECG's.

Another measure we used to assess the quality of the clustering was the Density Based Clustering Validation (DBCV) seen in Table 4 [Moulavi et al. 2014]. Overall, the DBCV scores were consistently high, with the exception of LBBB, which did not perform as well as the others, and AF, which performed poorly. For AF, this was expected due to the clear misclustering of diseased signals across the two cluster, as shown in Table 3. Although LBBB didn't perform poorly, it didn't achieve the same level of clustering quality as the other conditions, and we believe that it may present more complex patterns for sake of clustering.

Table 4. Clustering Quality Measures (DBCV).

Clustering	**ST**	**SB**	**AF**	**RBBB**	**LBBB**
F1 Score	0.9338	0.9146	0.7614	0.9173	0.9367
Density-Based Clustering Validation	0.9052	0.8874	0.1272	0.8938	0.6884

Our primary focus, however, is on the diseased ECG's that were misclassified into the non-diseased cluster. To further investigate, we compared these deviated ECG's with those that were accurately clustered. This comparison involved evaluating the discordance of cardiologists classifications against the original disease labels (see Table 5). The table reveals a significant increase in discordance for the deviated ECGs compared to those accurately clustered. The average discordance for the accurately clustered exams is 14.39%, while for the deviated exams, it rises to 38.98%, highlighting a pronounced difference in consistency between the two groups.

Table 5. Discordance rate between CODE-15% labels and senior cardiologist students classifications, for the diseased exams.

	ST	SB	AF	RBBB	LBBB
Accurately clustered Exams (%)	9.30	20.00	13.33	8.88	20.45
Deviated Exams (%)	50.00	51.33	11.76	40.91	40.91

By comparing the discordance rate of the misclustered ECGs with the average discordance rate among medicians, which typically ranges from 12%[Gomes et al. 2020], we see significant improvements, as shown in Table 6. This indicates that our algorithm is effectively identifying a select group of ECGs within a large dataset that are more likely to be challenging to diagnose. These cases, therefore, emerge as the most promising candidates for potential audit, underscoring the method's ability to prioritize cases that require greater clinical attention.

Table 6. Audit gain, between deviated diseased and random audit selection.

	ST	SB	AF	RBBB	LBBB	Average
Gain ratio (%)	316.0	325.0	-2.0	241.0	241.0	224.2

Overall, our findings suggest that the samples identified for audits are likely to be 224% more problematic than random sampling, and analyzing these cases could be particularly valuable in the context of medical audits. This result highlights the importance of prioritizing these misclustered ECG's in medical audits and discussions. By focusing on these cases, medical professionals can improve their assessments and potentially uncover additional information that may contribute to better patient outcomes and advances in medical knowledge.

6. Conclusion

In this paper, we performed five different clusterings to accurately separate normal and abnormal ECG's, aiming to minimize the number of misplaced diseased ECG's. The clusters for ST, SB, RBBB, and LBBB were formed with all F1 Scores above 0.9, while the AF cluster had the lowest F1 Score of 0.75, proving to be more challenging, indicating a need for improvement in future works.

The variation in which leads were best suited to each disease was also an interesting outcome. Although the best leads for diagnosing LBBB by doctors are D1, AVL, V1 and V6, they didn't perform very well for the computational distinction of this condition. In fact, a lead that isn't commonly employed for this diagnosis, V4, produced much better results. The reason for this is not yet clear, but it may indicate an unexplored potential for information in this lead.

Our approach, which targets ambiguous cases, has the potential to enhance diagnostic accuracy and efficiency by prioritizing ECGs that are more likely to present challenges or errors, ultimately improving the overall auditing process. Future work should focus on refining clustering techniques, expanding the range of conditions analyzed, increasing the number of validation samples, and exploring the diagnostic potential of less commonly used ECG leads.

7. Acknowledgments

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

References

Assent, I. (2012). Clustering high dimensional data. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2(4):340–350.

Banerjee, S. and Mitra, M. (2014). Application of cross wavelet transform for ecg pattern analysis and classification. IEEE Transactions on Instrumentation and Measurement, 63(2):326–333.

Berkaya, S. K., Uysal, A. K., Gunal, E. S., Ergin, S., Gunal, S., and Gulmezoglu, M. B. (2018). A survey on ecg analysis. Biomedical Signal Processing and Control, 43:216–235.

Bujang, M. A., Ab Ghani, P., Zolkepali, N. A., Adnan, T. H., Ali, M. M., Selvarajah, S., and Haniff, J. (2012). A comparison between convenience sampling versus systematic sampling in getting the true parameter in a population: Explore from a clinical database: The audit diabetes control management (adcm) registry in 2009. In 2012 international conference on statistics in science, business and engineering (ICSSBE), pages 1–5. IEEE.

Chiang, H.-T., Hsieh, Y.-Y., Fu, S.-W., Hung, K.-H., Tsao, Y., and Chien, S.-Y. (2019). Noise reduction in ecg signals using fully convolutional denoising autoencoders. IEEE Access, 7:60806–60813.

de Chazal, P., O'Dwyer, M., and Reilly, R. (2004). Automatic classification of heartbeats using ecg morphology and heartbeat interval features. IEEE Transactions on Biomedical Engineering, 51(7):1196–1206.

Fournier, Q. and Aloise, D. (2019). Empirical comparison between autoencoders and traditional dimensionality reduction methods. In 2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), pages 211–214. IEEE.

Gomes, P. R. et al. (2020). Sistema de laudos de eletrocardiograma: a importância de ferramentas de suporte à decisão.

HSE (2013). A Practical Guide to Clinical Audit. Health Service Executive (HSE) Quality and Patient Safety Directorate, Dublin, Ireland.

Johnston, G., Crombie, I., Alder, E., Davies, H., and Millard, A. (2000). Reviewing audit: barriers and facilitating factors for effective clinical audit. BMJ Quality & Safety, 9(1):23–36.

Kiranyaz, S., Ince, T., and Gabbouj, M. (2016). Real-time patient-specific ecg classification by 1-d convolutional neural networks. IEEE Transactions on Biomedical Engineering, 63(3):664–675.

Leland, M., John, H., and James, M. (2018). Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:1802.03426.

Lima, E. M., Ribeiro, A. H., Paixão, G. M. M., Ribeiro, M. H., Pinto-Filho, M. M., Gomes, P. R., Oliveira, D. M., Sabino, E. C., Duncan, B. B., Giatti, L., Barreto, S. M., Meira Jr, W., Schön, T. B., and Ribeiro, A. L. P. (2021). Deep neural network-estimated electrocardiographic age as a mortality predictor. Nature Communications, 12(1):5117.

Lyon, A., Mincholé, A., Martínez, J. P., Laguna, P., and Rodriguez, B. (2018). Computational techniques for ecg analysis and interpretation in light of their contribution to medical advances. Journal of The Royal Society Interface, 15(138):20170821.

Macfarlane, P., Devine, B., Latif, S., McLaughlin, S., Shoat, D., and Watts, M. (1990). Methodology of ecg interpretation in the glasgow program. Methods of Information in Medicine, 29(4):354 – 361. Cited by: 120.

Macfarlane, P. W. and Latif, S. (1996). Automated serial ecg comparison based on the minnesota code. Journal of Electrocardiology, 29:29–34.

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

Moulavi, D., Jaskowiak, P. A., Campello, R. J., Zimek, A., and Sander, J. (2014). Density-based clustering validation. In Proceedings of the 2014 SIAM international conference on data mining, pages 839–847. SIAM.

Nezamabadi, K., Sardaripour, N., Haghi, B., and Forouzanfar, M. (2023). Unsupervised ecg analysis: A review. IEEE Reviews in Biomedical Engineering, 16:208–224.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. Journal of machine learning research, 12(Oct):2825–2830.

Ribeiro, A. H., Paixao, G. M., Lima, E. M., Horta Ribeiro, M., Pinto Filho, M. M., Gomes, P. R., Oliveira, D. M., Meira Jr, W., Schon, T. B., and Ribeiro, A. L. P. (2021). CODE-15%: a large scale annotated dataset of 12-lead ECGs.

Ribeiro, A. H., Ribeiro, M. H., Paixão, G. M., Oliveira, D. M., Gomes, P. R., Canazart, J. A., Ferreira, M. P., Andersson, C. R., Macfarlane, P. W., Meira Jr, W., et al. (2020). Automatic diagnosis of the 12-lead ecg using a deep neural network. Nature communications, 11(1):1760.

Ribeiro, A. L. P., Paixão, G. M., Gomes, P. R., Ribeiro, M. H., Ribeiro, A. H., Canazart, J. A., Oliveira, D. M., Ferreira, M. P., Lima, E. M., de Moraes, J. L., Castro, N., Ribeiro, L. B., and Macfarlane, P. W. (2019). Tele-electrocardiography and big-data: The code (clinical outcomes in digital electrocardiography) study. Journal of Electrocardiology, 57:S75–S78.

Roopa, C. and Harish, B. (2017). A survey on various machine learning approaches for ecg analysis. International Journal of Computer Applications, 163(9):25–33.

Wang, W., Huang, Y., Wang, Y., and Wang, L. (2014). Generalized autoencoder: A neural network framework for dimensionality reduction. In 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 496–503.