Quantifying ECG Redundancy Through Mutual Information Analysis Among Leads and Its Application in CNNs

Elisa Ramírez[1], Samuel Ruipérez-Campillo[2], Raúl Alós[1], Francisco Castells[1], Rubén Casado-Arroyo[3], José Millet[1,4]

[1]ITACA Institute, Universitat Politècnica de València, Valencia, Spain
[2]Department of Computer Science, ETH Zurich, Switzerland
[3] Electrophysiology Lab, Hôpital Erasme, Brussels, Belgium
[4]Centro de Investigación Biomédica en Red Enfermedades Cardiovascular, Madrid, Spain.

Abstract

The electrocardiogram (ECG) is widely used in clinical practice for diagnosing diseases. However, conventional ECGs contain 12 projections of the three-dimensional cardiac dipole vector, resulting in significant redundancy which may negatively impact artificial intelligence algorithms performance. To address this issue, we developed a metric based on mutual information to quantify redundancy in standard ECGs. We used two strategies for redundancy reduction: eliminating leads and applying linear transformations to the original ECG leads. We employed a convolutional neural network (CNN) with inputs generated from redundancy elimination. We found that reducing the input to three orthogonal three-dimensional coordinates had minimal impact on the model's performance owing its slight decrease in performance to potential elimination of relevant information on the ECG wave morphologies. However, using six channels mitigated distortion by minimizing information loss, even with increased redundancy.

Clinical Relevance — *This study establishes an objective criterion for selecting cardiac vector projections that minimize redundancy while maintaining diagnostic capacity in the CNN. This serves as an objective and robust criterion for selecting the most informative projections in wearables or Holter monitors.*

1. Introduction

The conventional ECG recording represents the cardiac dipole in diverse orientations, captured through 12 non-orthogonal leads. These leads are categorized into standard (I, II, and III), augmented (aVR, aVL, aVF), and precordial (V1, V2, V3, V4, V5, V6) leads [1,2].

There are numerous studies that have employed Convolutional Neural Networks (CNNs) to automatically diagnose certain pathologies integrating data from all 12 ECG leads directly as input [3-5]. However, the inherent non-orthogonality of ECG leads results in the input data for these studies containing redundant information.

Furthermore, as the input size grows, the convolutional model's complexity also increases. Consequently, input redundancy becomes counterproductive, as it does not contribute new information and may lead to overfitting. So far, some studies have used lead elimination strategies without an objective criterion to ensure redundancy removal [6,7].

Quantifying this redundancy becomes crucial to objectively eliminate it. Redundancy reduction can be achieved, not only through lead elimination, but also via linear transformations that condense the 12-channel space into just three, representing coordinates in 3D space. Principal Component Analysis (PCA) and the Inverse Dower Transformation (IDT) [8] enable these transformations.

This study introduces a novel metric for quantifying redundancy in the input. This metric facilitates the identification of input configurations with minimal redundancy, thus allowing for the reduction of model complexity without compromising predictive capacity. As a direct consequence, this objective facilitates the identification of combinations that offer the highest informative capacity in scenarios with a limited number of leads, such as in Holter monitors or wearables.

2. Materials

We used the PTB-XL dataset accessible via Physionet [9], comprising 21,837 10-second and 12-lead ECG recordings from 18,885 patients. The dataset maintains an even distribution of genders (52% male, 48% female) and encompasses a broad age range (from 1 to 95 y.o.). It encompasses various heart pathologies, which have been annotated by two cardiologists as: normal (NORM), myocardial infarction (MI), ST/T changes (STTC), conduction disturbance (CD), and hypertrophy (HYP). The ECGs are sampled at both 500 Hz and 100 Hz.

To train the CNN, we divided the data into 88% for training and 12% for testing, ensuring that the same patient does not appear in both sets.

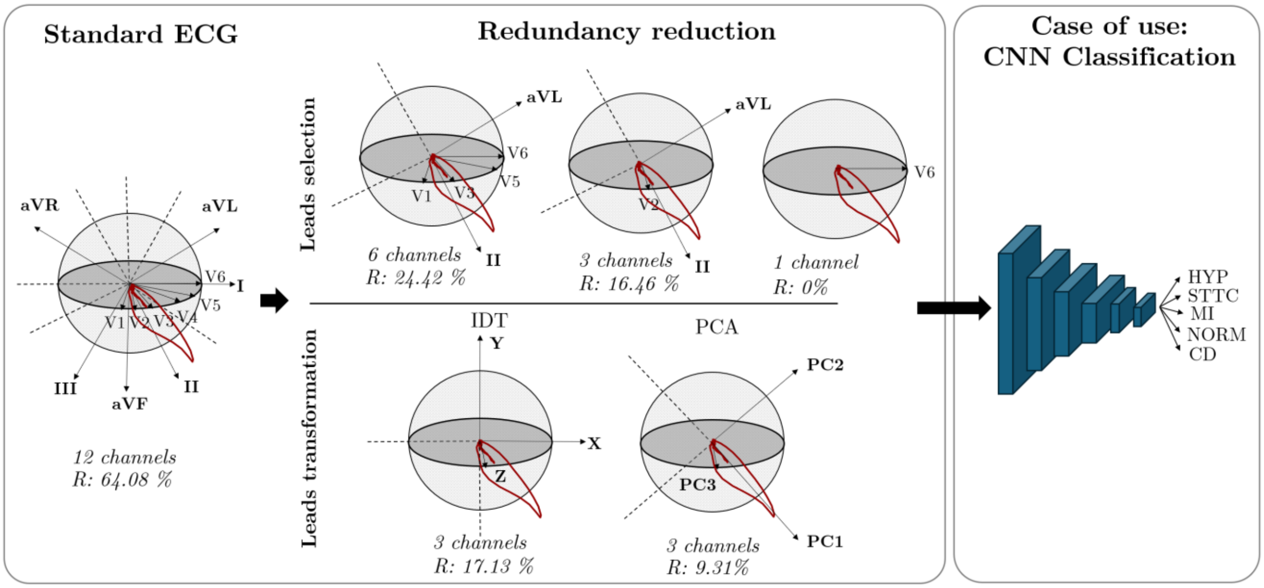

Figure 1. Reducing redundancy in the standard 12-lead ECG through lead selection and transformation strategies. A CNN was trained with these optimized inputs to classify five distinct pathologies.

3. Methods

3.1. Mutual Information between Projections

Treating the ECG signal as a sequence allows for the extraction of parameters from information theory, such as entropy, which quantifies the degree of uncertainty or dispersion of values within a specific ECG signal, defined as:

$$H(X) = -\sum_{x \in X} p(x) \ \log p(x) \qquad (1)$$

Furthermore, from entropy, the Mutual Information (I) between two independent variables (or ECG derivations) quantifies the dependency between them and can be calculated as:

$$I(X; Y) = H(X) + H(Y) - H(X,Y) \qquad (2)$$

The Normalized Mutual Information (NMI) can be computed by dividing the Mutual Information by the joint entropy.

3.2. Redundancy Score

3.2.1 Equation

Drawing from information theory, Redundancy (R) can be quantified as the subtraction of the conditional entropies to the joint entropy of the set as:

$$R(X_1, X_2, \dots, X_n) = \frac{H(X_1, X_2, \dots, X_n) - H(X_1|X_2, \dots, X_n) - \cdots - H(X_n|X_1, \dots, X_{n-1})}{H(X_1, X_2, \dots, X_n)} \qquad (3)$$

Figure 2.A graphically depicts the information in the datasets considered as Redundancy.

3.2.2 Redundancy Reduction Strategy

Leads selection

To acquire a set capable of capturing all the information of the cardiac dipole with minimal redundancy the most direct method consists of removing leads with high levels of shared information.

The optimal number would be three as it coincides with the minimum number to channels to represent coordinates in a three-dimensional space in which the cardiac dipole works. In this study, reductions to 3 leads (optimal number), 6 leads (half of the input and above the optimal) one lead (below the optimal) have been examined.

Leads transformation

The 12 leads correspond to projections of a three-dimensional vector. Therefore, with only three channels, all the information should be obtained without redundancy. To achieve this, the Inverse Dower Transformation (IDT) or Principal Component Analysis (PCA) can perform linear combinations of the 12 leads to obtain three coordinate axes in a three-dimensional space.

3.3. Case of Use in a CNN

Quantifying and reducing redundancy can be beneficial for optimizing the input of a CNN, aiming to simplify the model by providing non-redundant information and mitigating potential overfitting issues.

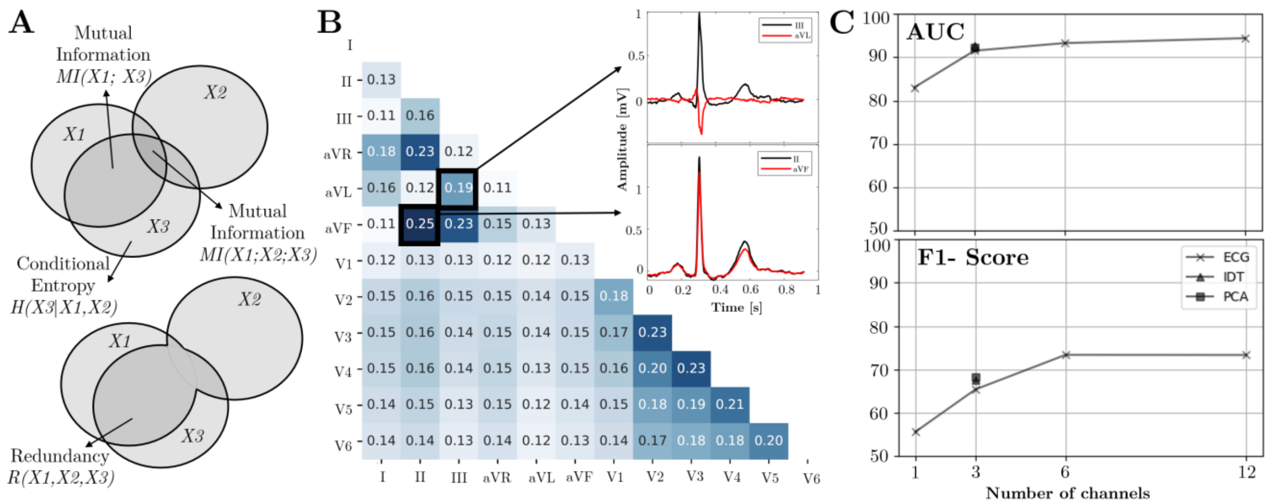

Figure 2. A) Graphical representation of information theory metrics (Conditional Entropy and Mutual Information) and the proposed Redundancy metric (see equation 3) for three sets: X1, X2 and X3. B) Normalized Mutual Information computed from equation 2 for all pairs of leads. C) Classification results for different number of inputs.

The selected architecture represents the best performing model as outlined in [3]. It comprises a 6-layer design integrating two-dimensional convolutional layers. The initial five layers are temporal, employing kernels with dimensions (1, N), where N denotes the kernel size. The final layer is spatial, utilizing kernels in the format (N, 1), aligning with the number of input channels. The ultimate layer of the model is dedicated to multiclass classification, addressing the five distinct pathologies.

The hyperparameters utilized for training the neural network mirror those reported in [3], as they were considered optimized through empirical search.

4. Results and Discussion

4.1. Mutual Information between Projections

Mutual Information captures shared information between two leads, and it accounts for similarities in morphology and amplitudes, as it operates on the probability distribution of signal values. Smaller angles between leads result in higher mutual information, and vice versa. However, differences are observed between pairs of leads with the same angle. For instance, II-aVF shows a NMI of 0.25, while III-aVL shows 0.19, despite both pairs having a 30-degree angle between them. This discrepancy arises because pairs with angles more similar to the direction of maximum variability (e.g., II-aVF) exhibit greater shared information than those orthogonal to the vector of maximum variability (e.g., III-aVL).

Considering this, relying solely on the angle between leads provides incomplete information about the shared information between them. Moreover, even though the theoretical angle between leads is known, imprecise electrode placement can result in slight changes in projection orientation, leading to variability in shared information.

Conversely, deterministic metrics such as Pearson Correlation Coefficient (PCC) quantify the similarity in morphology between two leads. However, this metric does not consider amplitude, meaning that two leads with significantly different orientations could still exhibit a high PCC.

Point-to-point errors, measured by metrics like Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), or Mean Squared Error (MSE), consider the signal's amplitude. However, they are dependent on the direction of projection, as, for example, a lead and its opposite with the same information captured might have very high error values.

4.2. Redundancy Score

The Redundancy metric quantifies the percentage of redundant information present in a specific set of channels. Both the lead selection strategy and the lead transformation strategy succeed in reducing the percentage of redundant information.

For lead selection, in the case of 6 leads, the selection was made by eliminating leads with high mutual information between them, as indicated by the matrix in Figure 2.B, resulting in the final set of leads: I, II, aVL, V1, V3, V5 and V6. For three leads, after testing various combinations, the least redundancy corresponded to leads II, aVL, and V2. In the case of a single lead, V6 corresponds to the most informative lead in terms of redundancy.

Reducing the number of leads from 12 to 6, and then to 3, results in a decrease in redundancy (see Table 1). When considering all designed inputs with three channels, PCA demonstrates the least redundancy for two primary reasons. Firstly, unlike IDT and 3-lead ECG, the channels in PCA are perfectly orthogonal. Secondly, one of the axes corresponds to the direction of maximum variability, while the other two are perpendicular, resulting in more independent variables, each representing variability along one of the three spatial axes.

4.3. Case of Use in a CNN

Reducing redundancy without eliminating information allows for a decrease in the input size of a CNN. As depicted in Figure 2.C, the predictive capacity of the neural network remains nearly constant until reaching three channels. As three ECG leads can never be perfectly orthogonal, transformations of ECG leads using IDT or PCA yield superior results.

The limitation encountered with three channels is the loss of information. In the case of a three-lead ECG, this is due to the imperfect orthogonality between channels. Additionally, in linear transformations, both approaches compute a single transformation from leads that are not obtained at the same angle, resulting in suboptimal representations of the cardiac dipole and the loss of certain information. With six leads, the loss of information due to non-orthogonality is mitigated by the excess of redundancy obtaining nearly the same results with half of the data volume. Conversely, with a single lead, the redundancy is 0, but due to significant information loss, the predictive performance decreases significantly. Yet, in future work could study whether this is enough for wearable systems and continuous-monitoring for patients in specific conditions.

Table 1. Model classification performance for different inputs by decreasing redundancy.

Input	Channels	R (%)	AUC	F1-Score
ECG	12	64.08	94.43	73.35
ECG	6	24.42	93.29	73.38
ECG	3	16.46	91.54	65.40
IDT	3	17.13	92.73	67.79
PCA	3	9.31	92.31	68.29
ECG	1	0	83.06	55.59

4. Conclusion

In this study, we presented a novel metric that quantifies the redundancy within a set of ECG leads. This has direct implications for simplifying convolutional models or other neural networks, thereby reducing training and prediction times and improving performance. Moreover, our study offers a robust criterion to select the most informative projections, particularly valuable for applications in wearables or Holter monitors.

Acknowledgments

This work was supported by PID2022-142514OB-I00 (National Research Program, Ministerio de Ciencia e Innovación, Spanish Government) and CIBERCV CB16/11/00486 (Instituto de Salud Carlos III).

References

[1] D. B. Geselowitz, "On the theory of the electrocardiogram," Proceedings of the IEEE, vol. 77, no. 6, pp. 857-876, 1989.

[2] Y. Sattar and L. Chhabra, "Electrocardiogram," in StatPearls. Treasure Island, FL: StatPearls Publishing, 2023. [Online].

[3] A. Anand, T. Kadian, M. K. Shetty, A. Gupta, "Explainable AI decision model for ECG data of cardiac disorders," Biomedical Signal Processing and Control, vol. 75, pp. 103584, 2022, ISSN 1746-8094, doi: 10.1016/j.bspc.2022.103584.

[4] U. B. Baloglu, M. Talo, O. Yildirim, R. S. Tan, U. R. Acharya, "Classification of myocardial infarction with multi-lead ECG signals and deep CNN," *Pattern Recognition Letters*, vol. 122, pp. 23-30, 2019, ISSN 0167-8655, doi: 10.1016/j.patrec.2019.02.016.

[5] A. Haque and N. Akhtar, "Ensemble model for ECG Classification," 2022 5th International Conference on Multimedia, Signal Processing and Communication Technologies (IMPACT), Aligarh, India, 2022, pp. 1-5, doi: 10.1109/IMPACT55510.2022.10029230.

[6] S. Jiménez-Serrano, M. Rodrigo, C. J. Calvo, J. Millet, F. Castells, "From 12 to 1 ECG lead: multiple cardiac condition detection mixing a hybrid machine learning approach with a one-versus-rest classification strategy," Physiological Measurement, vol. 43, no. 6, 2022, doi: 10.1088/1361-6579/ac72f5.

[7] Reyna MA, Sadr N, Perez Alday EA, Gu A, Shah A, Robichaux C, Rad BA, Elola A, Seyedi S, Ansari S, Li Q, Sharma A, Clifford GD. Will Two Do? Varying Dimensions in Electrocardiography: the PhysioNet/Computing in Cardiology Challenge 2021. Computing in Cardiology 2021;48: 1-4.

[8] Lars Edenbrandt, Olle Pahlm. Vectorcardiogram synthesized from a 12-lead ECG: Superiority of the inverse Dower matrix. Journal of Electrocardiology, vol. 21, no. 4, pp. 361-367, Apr. 1988. ISSN 0022-0736.

[9] Wagner, P., Strodthoff, N., Bousseljot, R., Samek, W., & Schaeffter, T. PTB-XL, a large publicly available electrocardiography dataset (version 1.0.3). PhysioNet, Jan, 2022.

Address for correspondence:
Elisa Ramírez
ITACA Institute Edificio 8G, Acceso B, 1a planta, Universitat Politècnica de València, Camino de Vera S/N 46022 Valencia España.
elracan@upv.edu.es