Does a Reduced ECG Lead Set Contain the Full 12-lead ECG Information for Interpretation?

Joel Xue[1,2], Trisha Dwivedi[1], Kim Barnett[1], Aleksandr Vlasenko[1], Miguel Kirsch[1], Aarya Parekh[1], Reena Yuan[1], Gari Clifford[2,4], David Albert[1]

[1]Alivecor Inc., Mountain View, CA, USA
[2]Emory University, Atlanta, GA, USA
[3]Georgia Institute of Technology, Atlanta, GA, USA

Abstract

Diagnostic 12-lead ECG analysis is one of the most important clinical tests in cardiology. However, a significant challenge with a 12-lead ECG is the need to attach multiple electrodes to obtain the signals. This study aims to use a deep learning (DNN)-based approach to directly generate interpretation results from the input waveform for a reduced lead set, and to compare the results with a DNN-based full 12-lead ECG model.

A DNN classification model was developed to detect selected morphology classes using 10-second ECG recordings with input lead sets that included the lead set of [I,II,V2,V4]. A second DNN model was built for the standard 12-lead set [I, II, V1-V6]. The training dataset consisted of two million 12-lead ECGs, and the test dataset included 500,000 ECGs, both from the Mayo Clinic.

The results indicated that reduced lead set exhibited similar excellent performance in detecting major morphology-related abnormalities, such as bundle branch block (AUC > 0.98), ischemia (AUC > 0.97), and myocardial infarction (old and acute, AUC > 0.98), as compared to the full 12-lead ECG. We conclude that a reduced ECG lead set can indeed contain sufficient information for full 12-lead ECG interpretation.

1. Introduction

Reduced ECG lead sets have been studied for years as alternatives to the standard 12-lead ECG, with the goal of simplifying ECG acquisition while maintaining diagnostic accuracy [1][2]. Various reduced lead sets have been evaluated, such as those comprising leads (I, II, V1, V5) and (I, II, V2, V5). Other studies have used non-standard electrode locations, such as a set with four chest leads and one reference lead [3]. The advantages of these lead systems include easier identification of lead positions, increased convenience for long-term monitoring of ischemia and infarction, and more. Although there have been several successful applications of reduced ECG lead sets, significant challenges remain before the medical community can accept them as replacements for the standard 12-lead ECG. These challenges include:

- The inability of a global lead conversion model to adjust to individual variations in body composition and movement during breathing.
- The reliance on standard 12-lead ECG criteria for interpretations.

In addressing these challenges, we are intrigued by a more fundamental question: Can a reduced ECG lead set provide the full information necessary for interpretation, with or without synthesized leads?

To answer this question, we employed deep learning neural networks (DNNs) to build an ECG interpretation algorithm directly from the ECG waveform, bypassing the need for explicit ECG feature detection. Recent studies have applied DNN models to ECG interpretation, where the original waveforms are ingested by the DNNs, using paired interpretation labels for training, without the need for explicit feature extraction or rule-based criteria [4,5,6,7]. While there has been some success with such approaches, it remains unclear whether the resulting DNNs might produce spurious results when applied to novel data.

Our approach bypasses the traditional steps of ECG feature detection and criteria-based interpretation, which is particularly relevant since no clinical criteria have been established for the reduced lead set. We hypothesize that a well-trained ECG interpretation DNN model can extract relevant information from the available leads to compensate for the missing lead data. Alternatively, we are curious about how much independent information might be lost when certain leads are excluded.

Aims of the Study: 1) To compare the interpretation performance of a full 12-lead DNN model with several reduced lead models, both with and without synthesized leads, and 2) To develop a working model for the clinical application of the selected reduced lead set.

An ECG interpretation algorithm typically includes two main components: rhythm interpretation and morphology interpretation. This study will focus on morphology interpretation, as it is more dependent on multiple precordial leads, which are crucial for detecting regional abnormalities.

2. Method

2.1. The Synthesized leads

A lead synthesis problem can generally be described as:

$$V_{syn} = F(W, Vx) \quad (1)$$

where:

- V_syn represents the vector of synthesized lead signals,
- Vx is the input vector of measured leads,
- W is the weight matrix, and
- F(·) represents a function with independent variables W and Vx, and dependent variables V_syn.

For linear models, the function F(·) takes the form of a linear equation (e.g., W×Vx), while for nonlinear models, F(·) represents a nonlinear function (e.g., a DNN model with multiple blocks of nonlinear activation functions).

During the training phase of the model, the error between the measured and synthesized leads is minimized, often using a least-squares term in the ensemble case for a global model. The final weight matrix W and/or function F(·) are determined once training is complete.

For the linear lead synthesis model in this study, we employed a least-squares method to minimize the difference between the estimated ECG and the actual measured ECGs.

2.2. Training and Validation Data

ECG data from 2.5 million samples were sourced from Mayo Clinic's Resting ECG database, with 2 million samples used for the training set and 500,000 independent ECGs from patients used for the validation set. All ECGs were labeled by Mayo Clinic cardiologists following their rigorous, high-quality processes.

2.3. ECG Signal Preprocessing

The original ECGs were 10 seconds long, recorded at 500 samples per second (sps). Since the models were designed for morphology interpretation, all ECG lead data were first filtered using a 40-Hz zero-phase shift low-pass filter and then downsampled to 150 sps. Compared to the 500 sps signal, the 150 sps data showed no degradation in model interpretation performance. The downsampled ECG lead data was used to form a median beat for each lead, which was then fed into the Deep Learning model.

2.4. Deep Learning Model Structure

The following three Convolutional Neural Network (CNN) deep learning models were built and trained separately with the following inputs:

1. **Full 12-lead inputs with 8 independent leads**: I, II, V1, V2, V3, V4, V5, V6.
2. **Reduced lead inputs only**: I, II, V2 (or V1), V4.
3. **Reduced lead inputs with synthesized leads**: I, II, V2, V3_syn, V4, V5_syn, V6_syn (where "_syn" denotes synthesized leads).

The input layer of the CNN model was designed to handle these three different input lead configurations.

The model structure is shown in Figure 1, and consists of:

a **Lead conversion layer**: This layer expands input limb leads I and II to generate leads III, aVR, aVL, and aVF. It then reorganizes them into the Cabrera format as: aVL, I, −aVR, II, aVF, III.

b. **Four residual-net blocks**: Each residual-net block combines a forward-pass convolution layer with the bypassed previous layer input, facilitating faster backpropagation learning [8].

c. **Three fully connected layers**: These layers classify embedded features from the CNN layers into the probabilities of the final determinations.

The inputs consist of median beats from 12-lead ECGs, which may include both measured and synthesized leads, depending on the selected reduced lead set. The output is the score from the final layer's activation function (sigmoid), representing the labeled ECG morphology interpretations, including:

- Right Bundle Branch Block (RBBB)
- Left Bundle Branch Block (LBBB)
- Left Ventricular Hypertrophy (LVH)
- Myocardial Infarctions (anterior, inferior, lateral, both previous and acute)
- Ischemia (anterior, inferior, lateral)
- Paced rhythm morphology
- Normal or Otherwise normal (defined as cases that do not meet the criteria for the above abnormal morphology classes).

2.5. Model Training

The model was trained with the following parameters:

- Batch size: 512
- Learning rate: 0.001
- Dropout rate: 0.05 for all dropout layers

To address the class imbalance problem common in multi-class classification models, a class-weight vector was applied during training to account for the varying prevalence of each class.[9]

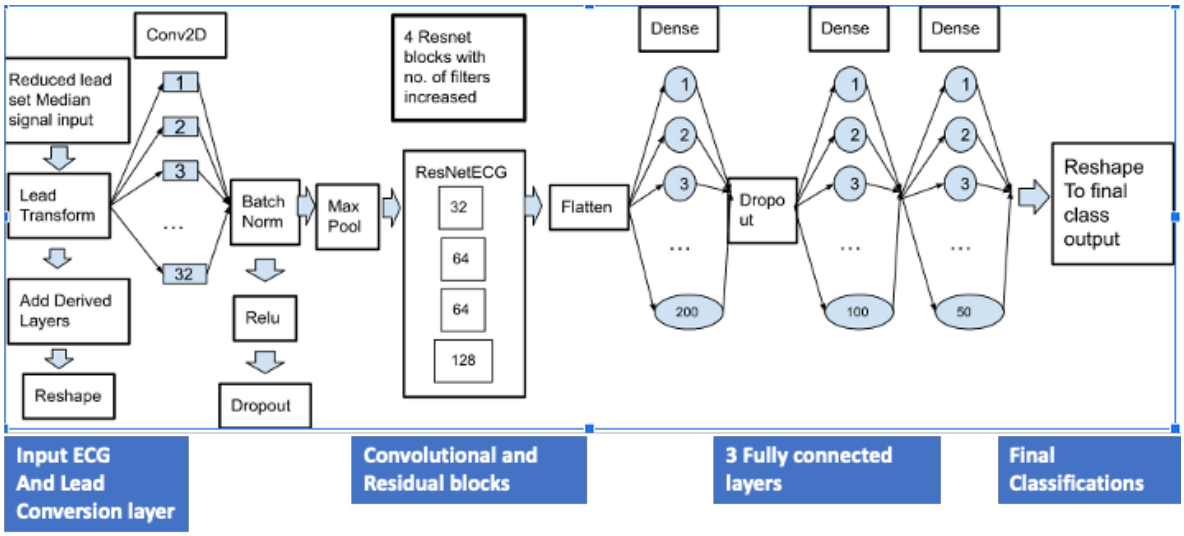

Figure 1. The CNN model for evaluating reduced lead sets' interpretation of morphology classes. The model includes a lead conversion layer, four residual-net blocks, and three fully connected layers. The three models trained in this study share the same structure, with differences only in the input conversion layer, which adapts to various input lead formats.

3. Results

The results, as shown in Table 1, demonstrate that the reduced lead set [I, II, V2, V4] performed comparably to the full 12-lead ECG in detecting major morphology-related abnormalities. For critical conditions such as bundle branch blocks (RBBB, AUCPR = 0.96; LBBB, AUCPR = 0.93) and myocardial infarction (Old MI, AUCPR = 0.77; Acute MI, AUCPR = 0.45), the reduced lead set achieved near-identical performance to the full 12-lead configuration.

The detection performance of left ventricular hypertrophy (LVH) was slightly lower with the reduced lead set (AUCPR = 0.65) compared to the full 12-lead ECG (AUCPR = 0.74). However, this difference did not significantly impact other classifications, indicating the robustness of the reduced lead set in maintaining diagnostic accuracy for most major morphology classes.

Notably, the addition of synthesized leads (V1_syn, V3_syn, V5_syn, V6_syn) to the reduced lead set did not yield any performance improvement. Classification metrics remained virtually unchanged across all categories, suggesting that the reduced lead set [I, II, V2, V4] alone is sufficient to carry equivalent information to that provided by the full 12-lead ECG for morphology interpretations.

Table 1: The Reduce lead set vs 12-lead set for the ECG morphology Interpretations

Morphology classes	Full 12-lead AUC	Full 12-lead AUCPR	Reduced lead [I, II, V2, V4] AUC.	Reduced lead [I, II, V2, V4] AUCPR	Reduced lead [I, II, V2, V4]+ Syn[v1, v3, v5, v6] AUC.	Reduced lead [I, II, V2, V4]+ Syn[v1, v3, v5, v6] AUCPR
RBBB	1.0	0.96	1.0	0.96	1.0	0.96
LBBB	1.0	0.93	1.0	0.93	1.0	0.93
LVH	0.98	0.74	0.97	0.65	0.97	0.65
Old MI	0.98	0.78	0.98	0.77	0.98	0.77
Acute MI	0.99	0.45	0.99	0.45	0.99	0.45
Pace	0.98	0.86	0.98	0.86	0.98	0.86
Normal	0.96	0.92	0.96	0.91	0.96	0.91

4. Discussions

This study investigated whether a reduced ECG lead set can provide sufficient information for accurate morphology interpretations compared to the standard 12-lead ECG. The results suggest that the reduced lead set [I, II, V2, V4] can perform equivalently to the full 12-lead ECG in identifying key morphological abnormalities, such as bundle branch blocks, myocardial infarctions, and ischemia. Furthermore, the addition of synthesized leads did not enhance the performance, highlighting that the core reduced lead set is already robust enough for reliable diagnostics.

Clinical Relevance of Reduced Lead Sets: ECG interpretation traditionally relies on a full 12-lead configuration to detect both rhythm and morphological abnormalities. This study focused specifically on morphology interpretation, which is more dependent on multiple precordial leads due to the regional specificity of morphological abnormalities like left ventricular hypertrophy (LVH) and myocardial infarctions. The results demonstrated that even with a reduced number of leads, essential diagnostic information can still be captured effectively.

One of the key findings is the comparable performance of the reduced lead set in detecting bundle branch blocks (RBBB and LBBB) and myocardial infarctions (both old and acute). The AUC/AUCPR values across these categories were virtually identical to those of the full 12-lead ECG, indicating that fewer leads can still provide comprehensive diagnostic insight for these common and clinically significant conditions.

Advantages of the Reduced Lead Approach: The reduced lead approach offers several practical benefits. First, it simplifies the ECG acquisition process by requiring fewer electrodes, making it more convenient for routine clinical settings and long-term monitoring. Additionally, the reduced complexity may increase patient comfort and compliance, especially during continuous monitoring scenarios.

The study also revealed that adding synthesized leads to the reduced set provided no measurable improvement in diagnostic accuracy. This suggests that the reduced lead set alone is sufficient for morphology interpretations, reinforcing the potential to streamline ECG setups without compromising clinical outcomes.

Limitations of Reduced and Synthesized Leads: Despite the promising results, there are limitations to using reduced and synthesized leads, particularly from the perspective of physician overreads. When interpreting ECGs with a reduced lead set, clinicians may face challenges in situations requiring more confirmation from those synthesized leads. For example, the interpretation of conditions like LVH, which was shown to have a lower AUPRC in the reduced lead set, may need more information from not directly measured leads.

In conclusion, this study supports the hypothesis that a reduced ECG lead set contains sufficient information for full 12-lead ECG interpretation, particularly for detecting major morphological abnormalities. The reduced lead set [I, II, V2, V4] performed equivalently to the full 12-lead configuration for most key diagnostic categories, with no added benefit from synthesized leads.

5. References

[1] S. P. Nelwan, J. A. Kors, S. H. Meij, J. H. Van Bemmel, and M. L. Simoons, "Reconstruction of the 12-lead electrocardiogram from reduced lead sets," *J Electrocardiol*, vol. 37, no. 1, pp. 11–18, 2004, doi: 10.1016/J.JELECTROCARD.2003.10.004.

[2] B. J. Drew, M. M. Pelter, D. E. Brodnick, A. V. Yadav, D. Dempel, and M. G. Adams, "Comparison of a new reduced lead set ECG with the standard ECG for diagnosing cardiac arrhythmias and myocardial ischemia," *J Electrocardiol*, vol. 35 Suppl, no. 4, pp. 13–21, 2002, doi: 10.1054/JELC.2002.37150.

[3] G. E. Dower, A. Yakush, S. B. Nazzal, R. V. Jutzy, and C. E. Ruiz, "Deriving the 12-lead electrocardiogram from four (EASI) electrodes," *J Electrocardiol*, vol. 21 Suppl, no. SUPPL., 1988, doi: 10.1016/0022-0736(88)90090-8.

[4] A. H. Ribeiro *et al.*, "Automatic diagnosis of the 12-lead ECG using a deep neural network," *Nature Communications 2020 11:1*, vol. 11, no. 1, pp. 1–9, Apr. 2020, doi: 10.1038/s41467-020-15432-4.

[5] B. Murugesan *et al.*, "ECGNet: Deep network for arrhythmia classification," in *MeMeA 2018 - 2018 IEEE International Symposium on Medical Measurements and Applications, Proceedings*, Institute of Electrical and Electronics Engineers Inc., Aug. 2018. doi: 10.1109/MeMeA.2018.8438739.

[6] A. H. Kashou, W.-Y. Ko, Z. I. Attia, M. S. Cohen, P. A. Friedman, and P. A. Noseworthy, "A comprehensive artificial intelligence–enabled electrocardiogram interpretation program," *Cardiovasc Digit Health J*, vol. 1, no. 2, pp. 62–70, Sep. 2020, doi: 10.1016/j.cvdhj.2020.08.005.

[7] J. Xue and L. Yu, "Applications of machine learning in ambulatory ECG," *Hearts*, vol. 2, no. 4, pp. 472–494, Oct. 2021, doi: 10.3390/hearts2040037.

[8] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. Accessed: Jul. 17, 2021. [Online]. Available: http://www.deeplearningbook.org

[9] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *J Big Data*, vol. 6, no. 1, Dec. 2019, doi: 10.1186/s40537-019-0192-5.

Address for correspondence:

Joel Xue
189 N Bernardo Ave
Mountain View, CA 94043
JXue656@Alivecor.com