Improving ECG Diagnosis of Left Ventricular Hypertrophy with Contrastive ECG-Echocardiogram Learning

Mously D Diaw[1,2], Stéphane Papelier[1], Alexandre Durand-Salmon[1], Julien Oster[2,3]

[1] Cardiabase, Banook Group, Nancy, France
[2] IADI, U1254, Inserm, Université de Lorraine, Nancy, France
[3] CIC-IT 1433, Université de Lorraine, Inserm, CHRU de Nancy, Nancy, France

Abstract

Because existing ECG criteria for left ventricular hypertrophy (LVH) have poor sensitivity at high specificity, we investigated whether deep learning (DL) could improve ECG-based LVH screening.

We trained a vision transformer via supervised contrastive ECG-Echocardiogram learning (CEEL), which uses, in addition to the cross-entropy (CE) loss, a loss function similar to the language-image CLIP loss to align ECG and Echo embeddings. The model, denoted ViT-CEEL, was trained on a MIMIC-IV subset of 2,028 patients. The same ViT architecture was also trained with the CE loss alone (ViT-CE). Both models were compared to traditional ECG-LVH methods (Cornell, Sokolow-Lyon, Romhilt-Estes) on a local study (AFICIONADO), which consisted of 57 asymptomatic patients but at high risk of heart failure.

On AFICIONADO, ViT-CE performs better (F_2 = 0.641, AUPRC = 0.38) than ViT-CEEL (F_2 = 0.556, AUPRC = 0.246), suggesting that CEEL does not boost performance as initially hypothesized. At similar specificity (0.941), the Cornell criterion has a better sensitivity (0.5) than ViT-CE (0.167). These results confirm previous findings that DL models based on the ECG alone achieve comparable performance to traditional ECG-LVH methods.

1. Introduction

Left ventricular hypertrophy (LVH) refers to a left-sided thickening of the heart muscle and is an independent predictor of cardiac dysfunction [1]. Its diagnosis is therefore important to the prevention of heart failure (HF), specially amongst patients at high risk of HF [2].

Because the ECG-LVH criteria used in clinical routine have poor sensitivity at high specificity [3], machine learning (ML), including deep learning (DL), has been proposed to improve the accuracy of ECG-based LVH diagnosis. However, the findings suggest that this is not achievable with the ECG alone but that the addition of clinical variables (e.g., age, sex, blood pressure) to the ML models could improve sensitivity [4–6].

There is yet to be an ECG-LVH method, independent of clinical variables, with both high sensitivity and specificity, leading experts to wonder if we have reached *the limits of the ECG to detect LVH* [7].

In this study, we hypothesized that training a DL model via contrastive ECG-Echo learning (CEEL) could improve sensitivity. Contrastive learning is a DL approach that allows to learn discriminative features from paired data. It has notably been successful in computer vision and natural language processing. CEEL uses a loss similar to the contrastive language-image pre-training (CLIP) loss proposed by Radford et al. [8]. It was inspired by Turgut et al. [9] who used the CLIP loss for contrastive ECG-CMR (Cardiac Magnetic Resonance) pre-training in order to estimate CMR parameters of cardiac function from the ECG. Our results do not suggest that CEEL improves LVH sensitivity and seem to confirm the previous findings that DL models based on the ECG alone achieve comparable performance to traditional ECG-LVH methods.

2. Material and methods

2.1. ECG/Echo data

Training set. The Medical Information Mart for Intensive Care (MIMIC)-IV database [10], available on PhysioNet [11], is a large database of patients admitted to intensive care units (ICUs) at the Beth Israel Deaconess Medical Center. From the MIMIC-IV-ECG module and MIMIC-IV-ECHO modules [12, 13], we identified 2,028 MIMIC-IV patients with a pair of 12-lead ECG and parasternal echo view. The delay between ECG recording and echo acquisition was typically $<$1 week and limited to 3 months. The 12-lead ECG recordings are 10 seconds in length and are sampled at 500 Hz. LV dimensions were extracted from the parasternal views using KerasOCR (Python li-

brary for text detection on images). LVM was calculated according to the Devereux formula [14]. LVM index (LVMI) was computed as the ratio between LVM and body surface area. LVH was subsequently defined as LVMI>95 g/m^2 for women and LVMI>115 g/m^2 for men [15].

Amongst the 2,028 patients, there were 971 men (48%) and 1,057 women, aged between 23 and 100 (median: 71, IQR: 62-81). Several HF risk factors were prevalent in this MIMIC-IV subset including essential (primary) hypertension (23.5%), hyperlipidemia (23%), nicotine dependence (17.7%) and obesity (11%). The median BMI was 28 kg/m^2 (IQR: 24.4-32.6) with 1,421 patients (70%) being overweight (BMI>25 kg/m^2). We identified 523 patients with LVH (prevalence of 26%), amongst which 288 women (55%).

Test set. The AFICIONADO study is an ongoing clinical trial (NCT06163911) at the Cardiology Department of the CHRU of Nancy, which aims to diagnose early stages of cardiac dysfunction from the ECG in an asymptomatic population at high risk of HF (with hypertension, dyslipidemia, obesity, etc.). So far, we identified 57 patients in this study with LVH labels, amongst which 33 men (58%). The patients are aged between 21 and 84 (median: 60, IQR: 47-68). The median BMI is 24.7 kg/m^2 (IQR: 22.1-28.1) with 32 patients (46%) being overweight. ECG and echo acquisitions were made the same day. The 12-lead ECG recordings are 10 seconds in length and are sampled at 1000 Hz. LVMI and LVH were defined similar to MIMIC-IV. 6/57 patients had LVH (prevalence of 10.5%), amongst which 5 women.

2.2. Deep learning

ViT-CEEL. The CEEL loss aims to maximize the similarity between matching ECG and echo pairs (ECG-Echo from the same patient) while minimizing the similarity between non-matching pairs. Denote x_1 the ECG modality, x_2 the Echo modality, f_1 the ECG encoder and f_2 the Echo encoder. In CEEL, the ECG features $f_1(x_1)$ and Echo features $f_2(x_2)$ are projected on a d-dimensional space through feature projectors g_1 and g_2. The resulting d-dimensional ECG and Echo embeddings are respectively

$$z_1 = g_1(f_1(x_1))$$

and

$$z_2 = g_2(f_2(x_2))$$

The CEEL loss is defined as

$$\text{CEEL} = \frac{1}{2}(L_1 + L_2) \tag{1}$$

where L_1 and L_2 are defined for the i^{th} ECG-Echo pair in a training batch of k pairs as

$$L_1 = -\log \frac{\exp(z_{1,i} \cdot z_{2,i}/\tau)}{\sum_{j=1}^{k} \exp(z_{1,j} \cdot z_{2,j}/\tau)}$$

and

$$L_2 = -\log \frac{\exp(z_{2,i} \cdot z_{1,i}/\tau)}{\sum_{j=1}^{k} \exp(z_{2,j} \cdot z_{1,j}/\tau)}$$

with dot product $z_1 \cdot z_2$ the measure of similarity between ECG and Echo embeddings (vice-versa for $z_2 \cdot z_1$). With L_1 (resp. L_2), the ECG (resp. Echo) in the i^{th} pair is compared with all Echos (resp. ECGs) in the batch. τ is the scaling parameter also called temperature.

Given the small size of the MIMIC-IV subset, we conducted supervised contrastive learning by summing the cross-entropy (CE) and CEEL losses (CE + CEEL).

For the ECG encoder, we used the vision transformer (ViT) of Turgut et al. [9], which takes as input 12-lead 5-second ECGs sampled at 500 Hz (ECGs in AFICIONADO were downsampled accordingly) and outputs 384 features. For the echo encoder, we used EchoNet-LVH, a DL model developed for the measurements of LV dimensions [16] that yields 2,048 echo features. The echo encoder and features projectors are not involved at test time. The resulting ViT is denoted ViT-CEEL.

ViT-CE. The same ViT arhitecture was also trained in the standard fully supervised (FS) fashion using the CE loss.

Training process. To train ViT-CEEL and ViT-CE, we randomly split the MIMIC-IV subset of 2,028 patients in 80%/20% training/validation sets. The models were trained using PyTorch (Adam optimizer). Training hyperparameters were finetuned by inspecting training/validation losses. The feature projectors output embeddings of size $d = 256$. The ECG feature projector g_1 is a feed-forward neural network (FNN) with an input layer of 384 neurons and an output layer of 256 neurons (no hidden layer). The Echo feature projector g_2 is a FNN of two hidden layers each having 2,048 neurons. The output layer has 256 neurons. For g_2, each layer in the FNN is followed by batch normalization and ReLU activation. The temperature parameter was fixed at $\tau = 0.07$.

ViT-CEEL and ViT-CE were trained during 200 and 500 epochs respectively. For ViT-CEEL, the initial learning rates (LRs) for the ECG encoder, the echo encoder and the feature projectors were 1e-5, 1e-5 and 0.1. For ViT-CE, the inital LR was 1e-3. Due to GPU memory limitations, we could only load 4 ECG-Echo pairs per batch for the training of ViT-CEEL, which was not ideal. To artificially increase the batch size (BS), we updated model weights every 64 mini-batches (gradient accumulation), which lead to an approximate BS of 4x64 = 256.

2.3. Traditional ECG criteria

We compared ViT-CEEL and ViT-CE to the Sokolow-Lyon criterion [17], the Cornell voltage criterion [18],

– **Sokolow-Lyon**: S_{V1} + (R_{V5} or R_{V6}) $\geq$ 3.5 mV

Table 1. LVH classification results on AFICIONADO.

Method	Accuracy	Sensitivity	Specificity	F_2	AUROC	AUPRC
Cornell	**0.93**	**0.5**	**0.98**	**0.536**	**0.768**	**0.516**
Sokolow-Lyon	0.895	0.333	0.961	0.357	0.598	0.242
Romhilt-Estes	0.877	0.5	0.922	0.484	0.824	0.251
ViT-CE	**0.807**	**0.833**	**0.804**	**0.641**	**0.85**	**0.38**
ViT-CEEL	0.825	0.667	0.843	0.556	0.755	0.246

– **Cornell**: R_{aVL} + $S_{V3} \geq$ 2.0 mV (women) or 2.8 mV (men)
and the **Romhilt-Estes** score [19]. ECG features were extracted from machine measurements (Mindray BeneHeart R12) associated with the ECG recordings in AFICIONADO.

3. Results

Table 1 provides the LVH classification results for the AFICIONADO study. AUROC denotes the Area Under the Receiver Operator Characteristic curve, and AUPRC, the Area under the Precision-Recall curve.

Amongst, the traditional ECG-LVH methods, the Cornell criterion yields the highest F_2 (0.536), AUROC (0.768) and AUPRC (0.516) scores. Amongst the DL models, ViT-CE yields the highest F_2 (0.641), AUROC (0.85) and AUPRC (0.38). The performance of ViT-CEEL is comparable to that of ViT-CE but CEEL does not seem to particularly boost performance. This might be due to the small batch size (4 ECG-Echo pairs per batch) chosen due to limited GPU memory.

At similar specificity (0.941), Cornell has a better sensitivity (0.5) than ViT-CE (0.167), which has the same sensitivity as the Romhilt-Estes score.

4. Discussion

These results confirm previous findings that both traditional and DL-based ECG-LVH methods yield poor sensitivity at high specificity. On AFICIONADO, the Cornell criterion seems to be the most sensitive method.

It is well known that the accuracy of ECG-based LVH diagnosis depends on several factors like LVH prevalence [20] and the underlying cause of LVH [21]. In our study, the Cornell criterion is more sensitive than the Sokolow-Lyon criterion likely because the AFICIONADO patients tend to be overweight [22]. Indeed, Sokolow-Lyon solely relies on precordial leads and might suffer from the attenuating effects of increased distance between the electrodes and the LV. The Romhilt-Estes score and the ViT models were developed on a different population than the AFICIONADO population—150 autopsied patients (60% LVH prevalence) for Romhilt-Estes—which could explain the poor performance of these methods on AFICIONADO.

DL usually requires a sufficiently large training set but the subset of 2,028 MIMIC-IV patients was the only ECG/Echo dataset available to us. A future access to databases like the UK Biobank [23], on which Turgut et al. [9] conducted their contrastive ECG-CMR pre-training, might allow to improve ViT-CEEL.

Because of the poor sensitivity of ECG-LVH methods, the American Heart Association [20] has suggested other research directions like the development of population-specific ECG-LVH criteria, for instance, by adapting the thresholds of voltage criteria, or risk stratification, which could be interesting to explore in the future.

5. Conclusion

LVH diagnosis is paramount to the prevention of adverse cardiovascular outcomes but ECG-based LVH diagnosis remains a great challenge. This work confirms the findings that, like traditional ECG-LVH criteria, DL models have poor sensitivity at high specificity when only relying on the ECG. Data scarcity is the main limitation of this study and better results might be obtained with CEEL by increasing the size of the training set and training batches.

Acknowledgments

This work was supported in part by Banook Group and in part by the French National Association for Research and Technology (ANRT, CIFRE) under Grant 2021/1466.

References

[1] Baker DW, Bahler RC, Finkelhor RS, Lauer MS. Screening for Left Ventricular Systolic Dysfunction among Patients with Risk Factors for Heart Failure. American Heart Journal 2003;146(4):736–740.

[2] Yancy CW, Jessup M, Bozkurt B, Butler J, Casey Jr DE, Drazner MH, Fonarow GC, Geraci SA, Horwich T, Januzzi JL, et al. 2013 ACCF/AHA Guideline for the Management of Heart Failure: Executive Summary: a Report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. Circulation 2013;128(16):1810–1852.

[3] Bressman M, Mazori AY, Shulman E, Chudow JJ, Goldberg Y, Fisher JD, Ferrick KJ, Garcia M, Di Biase L, Krumerman A. Determination of Sensitivity and Specificity of

Electrocardiography for Left Ventricular Hypertrophy in a Large, Diverse Patient Population. The American Journal of Medicine 2020;133(9):e495–e500.

[4] Sparapani R, Dabbouseh NM, Gutterman D, Zhang J, Chen H, Bluemke DA, Lima JA, Burke GL, Soliman EZ. Detection of Left Ventricular Hypertrophy Using Bayesian Additive Regression Trees: the MESA (Multi-Ethnic Study of Atherosclerosis). Journal of the American Heart Association 2019;8(5):e009959.

[5] Naderi H, Ramírez J, Van Duijvenboden S, Pujadas ER, Aung N, Wang L, Anwar Ahmed Chahal C, Lekadir K, Petersen SE, Munroe PB. Predicting Left Ventricular Hypertrophy from the 12-Lead Electrocardiogram in the UK Biobank Imaging Study Using Machine Learning. European Heart Journal Digital Health 2023;4(4):316–324.

[6] Kwon JM, Jeon KH, Kim HM, Kim MJ, Lim SM, Kim KH, Song PS, Park J, Choi RK, Oh BH. Comparing the Performance of Artificial Intelligence and Conventional Diagnosis Criteria for Detecting Left Ventricular Hypertrophy Using Electrocardiography. EP Europace 2020;22(3):412–419.

[7] Kashou AH, Noseworthy PA. Artificial Intelligence Capable of Detecting Left Ventricular Hypertrophy: Pushing the Limits of the Electrocardiogram?, 2020.

[8] Radford A, Kim JW, Hallacy C, Ramesh A, Goh G, Agarwal S, Sastry G, Askell A, Mishkin P, Clark J, et al. Learning Transferable Visual Models from Natural Language Supervision. In International Conference on Machine Learning. PMLR, 2021; 8748–8763.

[9] Turgut Ö, Müller P, Hager P, Shit S, Starck S, Menten MJ, Martens E, Rueckert D. Unlocking the Diagnostic Potential of ECG through Knowledge Transfer from Cardiac MRI. arXiv Preprint arXiv230805764 2023;.

[10] Johnson AE, Bulgarelli L, Shen L, Gayles A, Shammout A, Horng S, Pollard TJ, Hao S, Moody B, Gow B, et al. MIMIC-IV, a Freely Accessible Electronic Health Record Dataset. Scientific Data 2023;10(1):1.

[11] Goldberger A, Amaral L, Glass L, Hausdorff J, Ivanov PC, Mark R, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. Circulation 2000;101(23):e215–e220.

[12] Gow B, Pollard T, Nathanson LA, Johnson A, Moody B, Fernandes C, Greenbaum N, Berkowitz S, Moukheiber D, Eslami P, et al. MIMIC-IV-ECG: Diagnostic Electrocardiogram Matched Subset. PhysioNet 2023;.

[13] Gow B, Pollard T, Greenbaum N, Moody B, Johnson A, Herbst E, Waks JW, Eslami P, Chaudhari A, Carbonati T, et al. MIMIC-IV-ECHO: Echocardiogram Matched Subset. PhysioNet 2023;.

[14] Devereux RB, Alonso DR, Lutas EM, Gottlieb GJ, Campo E, Sachs I, Reichek N. Echocardiographic Assessment of Left Ventricular Hypertrophy: Comparison to Necropsy Findings. The American Journal of Cardiology 1986; 57(6):450–458.

[15] Marwick TH, Gillebert TC, Aurigemma G, Chirinos J, Derumeaux G, Galderisi M, Gottdiener J, Haluska B, Ofili E, Segers P, et al. Recommendations on the Use of Echocardiography in Adult Hypertension: a Report from the European Association of Cardiovascular Imaging (EACVI) and the American Society of Echocardiography (ASE). European Heart Journal Cardiovascular Imaging 2015;16(6):577–605.

[16] Duffy G, Cheng PP, Yuan N, He B, Kwan AC, Shun-Shin MJ, Alexander KM, Ebinger J, Lungren MP, Rader F, et al. High-Throughput Precision Phenotyping of Left Ventricular Hypertrophy with Cardiovascular Deep Learning. JAMA Cardiology 2022;7(4):386–395.

[17] Sokolow M, Lyon TP. The Ventricular Complex in Left Ventricular Hypertrophy as Obtained by Unipolar Precordial and Limb Leads. American Heart Journal 1949; 37(2):161–186.

[18] Casale PN, Devereux RB, Alonso DR, Campo E, Kligfield P. Improved Sex-Specific Criteria of Left Ventricular Hypertrophy for Clinical and Computer Interpretation of Electrocardiograms: Validation with Autopsy Findings. Circulation 1987;75(3):565–572.

[19] Romhilt DW, Estes Jr EH. A Point-Score System for the ECG Diagnosis of Left Ventricular Hypertrophy. American Heart Journal 1968;75(6):752–758.

[20] Hancock EW, Deal BJ, Mirvis DM, Okin P, Kligfield P, Gettes LS. AHA/ACCF/HRS Recommendations for the Standardization and Interpretation of the Electrocardiogram: Part V: Electrocardiogram Changes Associated with Cardiac Chamber Hypertrophy: a Scientific Statement from the American Heart Association Electrocardiography and Arrhythmias Committee, Council on Clinical Cardiology; the American College of Cardiology Foundation; and the Heart Rhythm Society: Endorsed by the International Society for Computerized Electrocardiology. Circulation 2009; 119(10):e251–e261.

[21] Hutchins SW, Murphy ML, Dinh H. Recent Progress in the Electrocardiographic Diagnosis of Ventricular Hypertrophy. Cardiology Clinics 1987;5(3):455–468.

[22] Abergel E, Tase M, Menard J, Chatellier G. Influence of Obesity on the Diagnostic Value of Electrocardiographic Criteria for Detecting Left Ventricular Hypertrophy. The American Journal of Cardiology 1996;77(9):739–744.

[23] Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, Downey P, Elliott P, Green J, Landray M, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLoS Medicine 2015;12(3):e1001779.

Address for correspondence:

Julien Oster (julien.oster@inserm.fr)
Laboratoire IADI (Inserm U1254)
Bâtiment Recherche, CHRU Nancy-Brabois, Rue du Morvan
54500 Vandoeuvre, France