The explainability of the latent variables is limited to the synthesis of electrocardiogram

Akinori Higaki ![ORCID] [1,2,]* and Osamu Yamaguchi ![ORCID] [2]

[1]Department of Intractable Disease and Aging Science, Ehime University Graduate School of Medicine, 454 Shitsukawa, Toon, Ehime 791-0295, Japan; and [2]Department of Cardiology, Pulmonology, Hypertension, and Nephrology, Ehime University Graduate School of Medicine, 454 Shitsukawa, Toon, Ehime 791-0295, Japan

We read with great interest the article by van de Leur et al.[1] reporting a method to extract interpretable features from electrocardiograms (ECGs) using a variational auto-encoder (VAE). In the present study, the authors used VAE to compress the ECG into 21 generative ECG factors (named FactorECG), which can be in turn used for predicting reduced ejection fraction (EF) and 1-year mortality. We commend the authors for presenting a new application of self-supervised machine learning by analyzing vast amounts of clinical data, as we are also working on the applications of autoencoders to nuclear cardiac imaging.[2] On the other hand, we are concerned that the title of the present study and the research methodology are somewhat misleading. Perhaps, there is a bit of confusion between interpretability and explainability involved.

Firstly, it is unclear what comparison is being made concerning the improvement in explainability. In this study, diagnostic performance is compared between the conventional deep convolutional neural network (DNN) model and the proposed method using VAE, but evidence of 'improved explainability' has not been provided. Only the correlation between classical ECG measurements and the proposed FactorECG is shown. It seems to us that authors are tying ECG measurements, which are inherently explainable, to more complicated and unexplainable parameters. For example, Factors 5, 10, and 25, which have been shown to be useful in diagnosing reduced EF using the extreme gradients boosting decision tree (XGBoost), reflect conventional ECG measurements, such as T-wave morphology, QT duration, and QRS width. If so, there is no need to use the circuitous pipeline but to directly train XGBoost with conventional ECG parameters. Then the Shapley Additive exPlanations values will ensure explainability. We think it would be more interesting to give up translating the FactorECG into conventional ECG indices and compare them with other clinical indices, such as echocardiographic parameters.

Secondly, the authors' criticism of the conventional DNN as a 'black box also misses the point. Their claim of 'model-level

explainability' is very limited to the generative process of ECG waveforms and does not directly explain reduced EF or 1-year mortality. By using the term 'pipeline', the authors seem to be intentionally hiding the fact that they are using logistic regression and XGBoost models for the downstream prediction tasks. It is also a factual error that the heat map only provides temporal locations of ECG features. We can find a handful of examples that important features are visualized on the ECG in the temporal and voltage directions at the same time.[3,4] It seems to us that the true strengths of their VAE-based approach are hampered by the issues described above.

We appreciate the importance of the present work in raising the interpretability (and explainability) of the generative model for synthesized ECG. The generation of realistic ECG by generative adversarial networks was recently reported by Thambawita et al.[5], but they were not able to control the morphology of synthesized ECG. Therefore, we hope van de Leur et al.[1] will continue to expand the use of generative models in ECG and cardiac imaging.

Data availability

The data underlying this article are available in the article.

References

1. van de Leur RR, Bos MN, Taha K, Sammani A, Yeung MW, van Duijvenboden S, Lambiase PD, Hassink RJ, van der Harst P, Doevendans PA, Gupta DK, van Es R. Improving explainability of deep neural network-based electrocardiogram interpretation using variational auto-encoders. *Eur Hear J* 2022;**3**:390–404.
2. Higaki A, Kawaguchi N, Kurokawa T, Okabe H, Kazatani T, Kido S, Aono T, Matsuda K, Tanaka Y, Hosokawa S, Kosaki T, Kawamura G, Shigematsu T, Kawada Y, Hiasa G, Yamada T, Okayama H. Content-based image retrieval for the diagnosis of myocardial perfusion imaging using a deep convolutional autoencoder. *J Nucl Cardiol* 2022;**8**. doi: 10.1007/s12350-022-03030-4.
3. Makimoto H, Höckmann M, Lin T, Glöckner D, Gerguri S, Clasen L, Schmidt J, Assadi-Schmidt A, Bejinariu A, Müller P, Angendohr S, Babady M, Brinkmeyer C, Makimoto A,

* Corresponding author. Tel: +81-89-960-5303, Fax: +81-89-960-5306, Email: keroplant83@gmail.com

Kelm M. Performance of a convolutional neural network derived from an ECG database in recognizing myocardial infarction. *Sci Rep* 2020;**10**:8445.

4. Rahman T, Akinbi A, Chowdhury MEH, Rashid TA, Şengür A, Khandakar A, Islam KR, Ismael AM. COV-ECGNET: COVID-19 detection using ECG trace images with deep convolutional neural network. *Heal Inf Sci Syst* 2022;**10**:1–16.

5. Thambawita V, Isaksen JL, Hicks SA, Ghouse J, Ahlberg G, Linneberg A, Grarup N, Ellervik C, Olesen MS, Hansen T, Graff C, Holstein-Rathlou N-H, Strümke I, Hammer HL, Maleckar MM, Halvorsen P, Riegler MA, Kanters JK. Deepfake electrocardiograms using generative adversarial networks are the beginning of the end for privacy issues in medicine. *Sci Rep* 2021;**11**:21896.