

Editorial



Machine Learning: a New Opportunity for Risk Prediction

Osung Kwon , MD¹, Wonjun Na , MA², and Young-Hak Kim , MD, PhD²

¹Division of Cardiology, Department of Internal Medicine, Eunpyeong St. Mary's Hospital, the Catholic University of Korea, Seoul, Korea

²Division of Cardiology, Department of Internal Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea

OPEN ACCESS

Received: Sep 26, 2019

Accepted: Oct 7, 2019

Correspondence to

Young-Hak Kim, MD, PhD

Division of Cardiology, Department of Internal Medicine, Asan Medical Center, University of Ulsan College of Medicine, 88, Olympic-ro 43-gil, Songpa-gu, Seoul 05505, Korea.

E-mail: mdyhkim@amc.seoul.kr


Copyright © 2020. The Korean Society of Cardiology

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original work is properly cited.


ORCID iDs

Osung Kwon 

<https://orcid.org/0000-0001-6857-8632>

Wonjun Na 

<https://orcid.org/0000-0003-1660-8101>

Young-Hak Kim 

<https://orcid.org/0000-0002-3610-486X>

Conflict of Interest

The authors have no financial conflicts of interest.

Author Contributions

Conceptualization: Kwon O, Kim YH; Writing - original draft: Kwon O, Na W; Writing - review & editing: Kim YH.

► See the article "Development and External Validation of a Deep Learning Algorithm for Prognostication of Cardiovascular Outcomes" in volume 50 on page 72.

Globally, cardiovascular disease (CVD) remains the major cause of mortality and morbidity, and identifying people at risk of CVD is the cornerstone of clinical cardiology.¹ Accordingly, current guidelines for primary prevention of CVD recommend algorithms to identify asymptomatic patients on the basis of their predicted risk.^{2,3} These established algorithms are typically developed using multivariate regression models with a limited number of well-established risk factors and generally assume that all such factors are related to the CVD outcomes in a linear fashion, with limited or no interactions between the different factors.^{4,5} Owing to their restrictive modeling assumptions and limited number of predictors, the existing algorithms generally exhibit suboptimal predictive performance.^{4,5}

Along with the emergence of big data, machine learning (ML) provides an alternative approach to established prediction modelling that may address the current limitations. Accumulated medical data and digitalized clinical information enable ML to verify a hypothesis generated from a conventional statistical analysis and to agnostically discover new predictors of CVD risk. Recently, deep learning (DL), a branch of ML, has become increasingly popular in the medical research community because of its excellent performance in different domains and the rapid methodological improvements.⁶ DL represents an improvement in artificial neural networks, consisting of more layers that permit higher levels of abstraction and improved predictions from data.⁷ To date, it is the leading ML tool in medical image analysis, with promising results.⁸ By virtue of large training biomedical data and advanced computing power, more recently, DL has been applied to the development of risk prediction models using electronic health data.⁶

Based on this background, Cho et al. investigated the additional discriminative accuracy of a time-series DL algorithm using repeated-measures data for identifying people at high risk of CVD, in comparison with the Cox hazard regression model.⁹ The authors found that the time-series DL algorithm analysis showed greater discriminative accuracy than the Cox model approaches. This study expands the possibility of DL from models that predict outcomes on the basis of data from specific time points to those that predict future events in complex time-varying datasets. The study used large data from a national health screening program and the national health insurance claims database in South Korea for development

The contents of the report are the author's own views and do not necessarily reflect the views of the *Korean Circulation Journal*.

and validation. Furthermore, prospective cohort data from the Rotterdam Study were used for ethnically generalizable external validation. The design based on this approach abided by the recommendation of the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis statement,¹⁰⁾ which made the study results reliable. In addition, ML, particularly neural networks, is sometimes called “black box” because of the difficulty of interpretation. Thus, the authors assessed the attribute rank of risk predictors in the DL model.

The data derived from the hospital information system are characterized by tremendousness, heterogeneity, and complexity, and DL could provide a solution for analyzing these kind of complex data. However, as mentioned by the authors in the limitations, in the study, only 6 variables that are already well-known strong cardiovascular risk factors, were used to develop the risk prediction model, which limits the study value and use of DL. Considering the purpose of the study, which was to confirm the superior analytic performance of DL in contrast to that of Cox regression, the study serves as an attempt to navigate the challenges of DL for developing CV risk prediction models. Further studies using a large number of diverse variables would be required to validate the predictive performance of DL.

DL introduces exciting new opportunities for precision medicine, including risk stratification and future event prediction. Attempts to apply DL methods to patient care and clinical research are already planned or underway. In spite of the current hurdles to the application of DL for cardiac risk stratification, inspiration and dispassionate effort would lead to the development of more reliable and robust models for realizing personalized cardiovascular health care.

REFERENCES

1. Roth GA, Johnson C, Abajobir A, et al. Global, regional, and national burden of cardiovascular diseases for 10 causes, 1990 to 2015. *J Am Coll Cardiol* 2017;70:1-25.
[PUBMED](#) | [CROSSREF](#)
2. Goff DC Jr, Lloyd-Jones DM, Bennett G, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association task force on practice guidelines. *J Am Coll Cardiol* 2014;63:2935-59.
[PUBMED](#) | [CROSSREF](#)
3. Authors/Task Force Members Piepoli MF, Hoes AW, et al. 2016 European Guidelines on cardiovascular disease prevention in clinical practice: the sixth joint task force of the European Society of Cardiology and other societies on cardiovascular disease prevention in clinical practice (constituted by representatives of 10 societies and by invited experts) developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR). *Atherosclerosis* 2016;252:207-74.
[PUBMED](#) | [CROSSREF](#)
4. Siontis GC, Tzoulaki I, Siontis KC, Ioannidis JP. Comparisons of established risk prediction models for cardiovascular disease: systematic review. *BMJ* 2012;344:e3318.
[PUBMED](#) | [CROSSREF](#)
5. Alaa AM, Bolton T, Di Angelantonio E, Rudd JH, van der Schaar M. Cardiovascular disease risk prediction using automated machine learning: a prospective study of 423,604 UK Biobank participants. *PLoS One* 2019;14:e0213653.
[PUBMED](#) | [CROSSREF](#)
6. Miotto R, Wang F, Wang S, Jiang X, Dudley JT. Deep learning for healthcare: review, opportunities and challenges. *Brief Bioinform* 2018;19:1236-46.
[PUBMED](#) | [CROSSREF](#)
7. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521:436-44.
[PUBMED](#) | [CROSSREF](#)

8. Litjens G, Ciompi F, Wolterink JM, et al. State-of-the-art deep learning in cardiovascular image analysis. *JACC Cardiovasc Imaging* 2019;12:1549-65.
[PUBMED](#) | [CROSSREF](#)
9. Cho IJ, Sung JM, Kim HC, et al. Development and external validation of a deep learning algorithm for prognostication of cardiovascular outcomes. *Korean Circ J* 2019;50:72-84.
[PUBMED](#) | [CROSSREF](#)
10. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 2015;162:55-63.
[PUBMED](#) | [CROSSREF](#)