Title

Deep Neural Network-Based Prediction of the Risk of Advanced Colorectal Neoplasia.

Journal

Gut and Liver, 15(1)

Authors

Min, Jun

Yang, Hyo-Joon

Kwak, Min

et al.

Publication Date

2021-01-15

DOI

10.5009/gnl19334

Peer reviewed

Gut and Liver
https://doi.org/10.5009/gnl19334
pISSN 1976-2283 eISSN 2005-1212

Original Article

Deep Neural Network-Based Prediction of the Risk of Advanced Colorectal Neoplasia

Jun Ki Min[1], Hyo-Joon Yang[2], Min Seob Kwak[1], Chang Woo Cho[3], Sangsoo Kim[3], Kwang-Sung Ahn[4], Soo-Kyung Park[2], Jae Myung Cha[1], and Dong Il Park[2]

[1]Department of Internal Medicine, Kyung Hee University Hospital at Gangdong, Kyung Hee University School of Medicine, [2]Division of Gastroenterology, Department of Internal Medicine and Gastrointestinal Cancer Center, Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, [3]Department of Bioinformatics, Soongsil University, and [4]Functional Genome Institute, PDXen Biosystems Inc., Seoul, Korea

Article Info

Received September 30, 2019
Revised December 6, 2019
Accepted December 9, 2019
Published online December 31, 2020

Corresponding Author

Jae Myung Cha
ORCID https://orcid.org/0000-0001-9403-230X
E-mail drcha@khu.ac.kr

Dong Il Park
ORCID https://orcid.org/0000-0003-2307-8575
E-mail diksmc.park@samsung.com

Jun Ki Min and Hyo-Joon Yang contributed equally to this work as first authors.

Background/Aims: Risk prediction models using a deep neural network (DNN) have not been reported to predict the risk of advanced colorectal neoplasia (ACRN). The aim of this study was to compare DNN models with simple clinical score models to predict the risk of ACRN in colorectal cancer screening.

Methods: Databases of screening colonoscopy from Kangbuk Samsung Hospital (n=121,794) and Kyung Hee University Hospital at Gangdong (n=3,728) were used to develop DNN-based prediction models. Two DNN models, the Asian-Pacific Colorectal Screening (APCS) model and the Korean Colorectal Screening (KCS) model, were developed and compared with two simple score models using logistic regression methods to predict the risk of ACRN. The areas under the receiver operating characteristic curves (AUCs) of the models were compared in internal and external validation databases.

Results: In the internal validation set, the AUCs of DNN model 1 and the APCS score model were 0.713 and 0.662 ($p<0.001$), respectively, and the AUCs of DNN model 2 and the KCS score model were 0.730 and 0.667 ($p<0.001$), respectively. However, in the external validation set, the prediction performances were not significantly different between the two DNN models and the corresponding APCS and KCS score models (both $p>0.1$).

Conclusions: Simple score models for the risk prediction of ACRN are as useful as DNN-based models when input variables are limited. However, further studies on this issue are warranted to predict the risk of ACRN in colorectal cancer screening because DNN-based models are currently under improvement. **(Gut Liver 2021;15:85-91)**

Key Words: Colorectal neoplasms; Deep learning; Neural networks; Prediction; Mass screening

INTRODUCTION

Colorectal cancer (CRC) is one of the major cancers whose incidence is steadily increasing in many countries, including Korea.[1] CRC screening is able to reduce CRC-related mortality and morbidity,[2,3] but, challenged by limited resources and low adherence.[4] Therefore, risk prediction model to predict the risk of advanced colorectal neoplasia (ACRN) may improve the effectiveness of CRC screening. This strategy was developed to identify individuals who are at high risk of ACRN, and judiciously to use the limited resources of colonoscopy for the high-risk population rather than in low-risk population. Recently, risk stratification models, such as the Asian-Pacific Colorectal Screening (APCS) score model, increased the effectiveness of CRC screening.[5-10] However, simple score models were limited as they used logistic regression (LR) models,[5-10] which have low sensitivity and high false positivity because of the limited variables and performance levels of the LR method.

Deep learning model using deep neural network (DNN) is computational models composed of multiple processing layers to learn the representations of the data with multiple

levels of abstraction.[11,12] DNN techniques have reported to improve the diagnostic accuracy in the diagnosis of skin cancer,[13] diabetic retinopathy,[14,15] lymph node metastasis of breast cancer,[16] and colorectal adenomas during colonoscopy.[17,18] Furthermore, DNN techniques may provide better risk prediction models for the ACRN detection as they can utilize clinical data more efficiently than previous LR models. However, no DNN-based risk prediction model was reported to predict the risk of ACRN. DNN-based risk prediction models may provide better predictive power. Although simple score models have the advantage of easy-to-use in daily clinical practice. they were limited by the lack of external validation,[5,7,10] which is important in terms of overfitting.

This study was aimed to compare the performances of DNN-based risk prediction models with those of simple score models (i.e., LR models) to predict the risk of ACRN.

MATERIALS AND METHODS

1. Study population

The database of screening colonoscopy at Health Screening Center of Kangbuk Samsung Hospital (cohort 1, n=121,794) between January 2003 and December 2012 was used as a training, tuning, and internal validation set (Fig. 1).[10] The database of screening colonoscopy from Kyung Hee University Hospital at Gangdong between September 2006 and September 2009 (cohort 2, n=3,738)[19] was also used as an external validation set to prevent a bias of an input data from a single hospital. Overall, 51,458 subjects were excluded from cohort 1 (Fig. 1A) and 409 subjects were excluded from cohort 2 (Fig. 1B) with the same exclusion criteria: history of previous colorectal examinations such as barium enema, sigmoidoscopy, or colonoscopy, history of CRC or inflammatory bowel disease, history of colorectal surgery, incomplete colonoscopy due to cecal intubation failure or inadequate bowel preparation, and missing clinical data. As a result, 70,336 subjects from cohort 1 were randomized in a ratio of 7:1:2 into a training set (n=49,235), tuning set (n=7,034), and internal validation set (n=14,067), whereas the data of 3,561 subjects from cohort 2 were used for the external validation set. Two DNN models were developed and compared their performances to predict the risk of ACRN with APCS and Korean Colorectal Screening (KCS) score models. This retrospective study was approved by the Institutional Review Board of both institutions (IRB numbers: 2017-07-02 for Kangbuk Samsung Hospital and KHNMC 2019-12-004 for Kyung Hee University Hospital). The informed consent was waived because of the retrospective design and anonymized patient data.

2. Database

The demographic data, body mass index (BMI), smoking status, family history of CRC in a first-degree relative, colonoscopy findings, and pathology reports were reviewed by a physician or a specially trained, non-physician research nurse as described previously.[10,20] A current smoker was defined as one who consumed at least one pack of cigarettes per week. Positive family history of CRC was defined as positive CRC history in at least one first-degree relative. According to the Asian-Pacific guidelines, BMI ≥25 kg/m^2 was defined as obesity.[21] The input variables included age, sex, family history of CRC, smoking, and BMI, and the output (labelled) data were collected from the colonoscopy reports and pathology results. APCS score model used four variables: age (<50, 50–60, 60–70, and ≥70

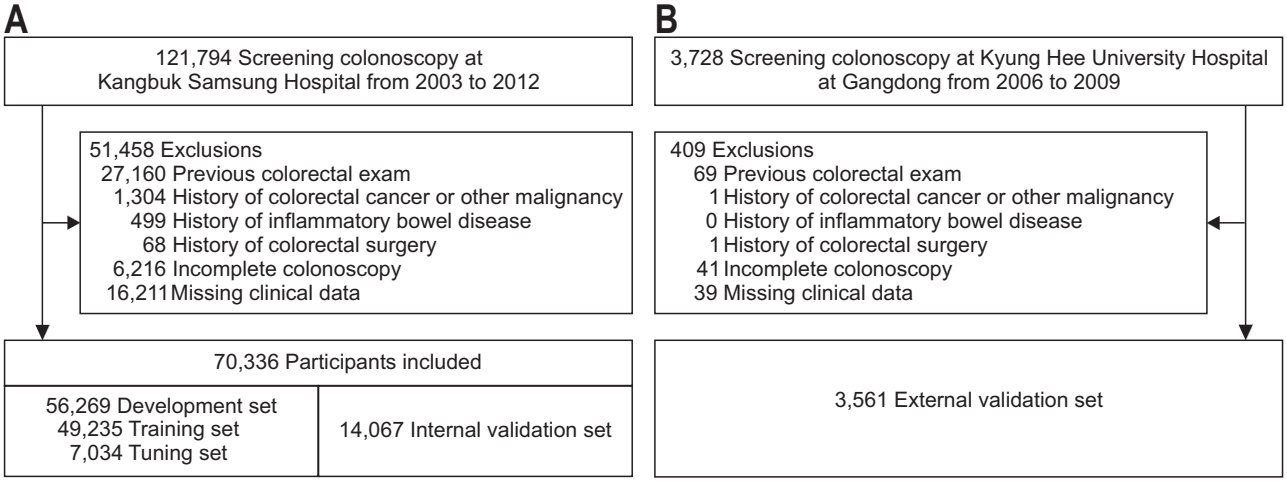

Fig.1. Flowchart of the inclusion and exclusion of study populations. (A) Kangbuk Samsung Hospital and (B) Kyung Hee University Hospital at Gangdong.

years), sex, smoking (none/past and current), and family history of CRC,[6] whereas KCS score model used five variables: age, sex, smoking, family history of CRC, and BMI.[19]

3. Colonoscopy protocol

The board-certified endoscopists performed all colonoscopies using Evis Lucera CV-260 colonoscopes (Olympus Medical Systems, Tokyo, Japan) in cohort 1 and using EG-590WR colonoscopes (Fujinon Inc., Saitama, Japan) in cohort 2. Bowel preparation was performed with 4 L of polyethylene glycol solution in both hospitals. All the polyps were measured for their size and were removed by a biopsy or polypectomy. The histological specimens were evaluated by gastrointestinal pathologists. ACRN was defined as a colorectal carcinoma or an advanced adenoma (any adenoma ≥1 cm in size, or with villous component or high-grade dysplasia).[10]

4. Development of DNN

LR models were fitted to the training set as a comparator for APCS and KCS score models (Table 1). As the DNN framework, a feedforward neural network[22] as the DNN structure and Google's TensorFlow (version 1.4.1)[23] in Python (version 2.7.6.) were used. Two DNNs were developed: DNN model 1 used the four variables of the APCS score model and DNN model 2 used the five variables of the KCS score model. All continuous variables were standardized for feature scaling.[24] The training set was used for model learning and the tuning set served as hyperparameter tuning to avoid overfitting. Both DNNs had two hidden layers and seven and eight nodes for each layer based on experiments involving different hyperparameters (Fig. 2). The DNNs used Adam[25] as the optimizer with learning rate of 0.1, the Xavier initializer[26] to initialize the weights of hidden units, and the exponential linear unit activation function in each layer.[27] As proposed by Kingma and Ba,[25] Adam was used as an optimization algorithm with β_1=0.9, β_2=0.999, and $\varepsilon=10^{-8}$. Dropout was applied after the activation function in the last hidden layer to prevent overfitting,[28] and the softmax function linked the final hidden layer to the output layer. Each model was trained for 1,000 epochs using the same training set. The output values generated from the trained networks demonstrated the probability for each input case with ACRN, where the range of

Table 1. Baseline Characteristics of the Cohorts and the APCS and KCS Scores

Covariates	APCS score			KCS score		
	β	OR (95% CI)	p-value	β	OR (95% CI)	p-value
Age group, yr			<0.001			<0.001
<50	1	1		1	1	
50–69	1.566	4.79 (4.10–5.60)		1.561	4.76 (4.07–5.57)	
≥70	2.255	9.54 (5.66–16.08)		2.271	9.68 (5.74–16.33)	
Male sex	0.548	1.73 (1.44–2.07)	<0.001	0.488	1.63 (1.36–1.96)	<0.001
Current or past smoker	0.315	1.37 (1.17–1.61)	<0.001	0.315	1.37 (1.17–1.61)	<0.001
Family history of CRC	0.030	1.03 (0.87–1.22)	0.734	0.045	1.05 (0.74–1.47)	0.796
BMI ≥25 kg/m^2	-	-	-	0.314	1.37 (1.17–1.60)	<0.001
Constant	–5.199	0.01 (0.00–0.01)	<0.001	–5.275	0.01 (0.00–0.01)	<0.001

APCS, Asian-Pacific Colorectal Screening; KCS, Korean Colorectal Screening; OR, odds ratio; CI, confidence interval; CRC, colorectal cancer; BMI, body mass index.

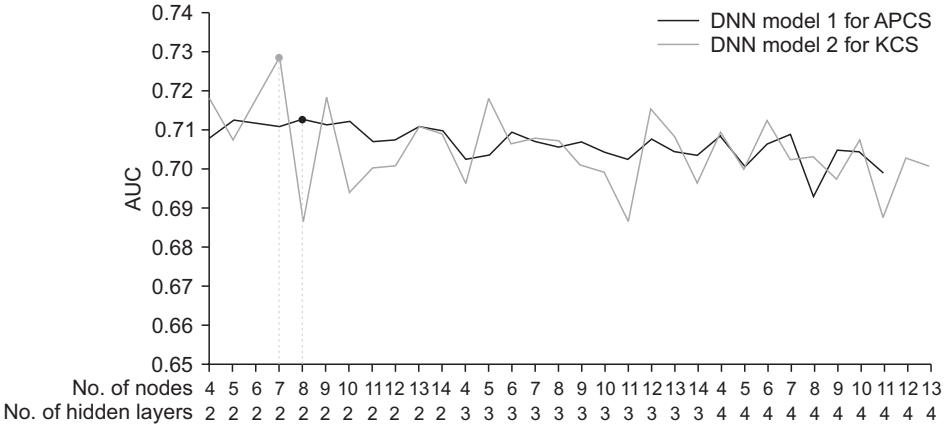

Fig. 2. Comparison of the performances of the deep neural network (DNN) models with different hyperparameter values. All DNNs presented in the table used the exponential linear unit activation function, Xavier initializer, Dropout for normalization, and Adam optimizer with 1,000 epochs of the same training set for each model.
AUC, area under the receiver operating characteristic curve; APCS, Asian-Pacific Colorectal Screening; KCS, Korean Colorectal Screening.

output was between low (0) and high (1) probability.

5. Statistical analysis

The primary outcome was comparison of the performances of the DNN models against the LR models to predict the risk of ACRN in the external validation set. The area under the curve (AUC) of the receiver operating characteristic curve of each model was compared with that of the others using the DeLong test.[29] The AUC was 0.68 and the prevalence of ACRN was 1.4% in our previous study with a LR model.[10] With the assumption that an increment of at least 0.05 in the AUC of DNN models will be clinically significant, a minimum sample size of 13,064 was required for statistical power of 80%, p<0.05 level of significance, and strong correlation (correlation coefficient, 0.7) between the models, both in the positive and negative cases.[30] R statistical program, version 3.3.2 (R Development Core Team, Vienna, Austria), was used for statistical analyses. All p-values were two-sided, and p<0.05 was considered statistically significant.

RESULTS

1. Baseline characteristics

The baseline characteristics of the subjects of both cohorts have been described in previous reports.[10,19] In cohort 1, the mean age was 41.6±8.3 years and 48,810 patients were male (69.4%). Of the 10,620 subjects ≥50 years old (15.1%), 414 subjects had ACRN (3.9%). There were no significant differences in the demographic and clinical data between the training, tuning, and internal validation sets (Table 2). In cohort 2, the mean age was 51.3±9.0 years and 2,152 patients were male (60.4%). Of the 2,048 subjects ≥50 years old (57.5%), 146 subjects had ACRN (7.1%). The subjects of the external validation set were relatively older and less male dominant, and had higher rate of ACRN and smokers than those of internal validation set.

2. Performance of DNN models

The receiver operating characteristic curves of the APCS score and DNN model 1 for internal and external validation set are illustrated in Fig. 3. When compared with APCS score model (AUC, 0.662; 95% confidence interval [CI], 0.619 to 0.705) in the internal validation set, DNN model 1 showed a good discrimination with a significantly improved prediction performance (AUC, 0.713; 95% CI, 0.674 to 0.752; p<0.001) (Fig. 3A).[31] On the contrary, DNN model 1 in the external validation set failed to show performance improvement (AUC, 0.754; 95% CI, 0.719 to 0.790) than those of APCS score model (AUC, 0.742; 95% CI, 0.707 to 0.777) (p=0.433) (Fig. 3B). The comparison of the performance of the KCS score model and DNN model 2 are illustrated in Fig. 4. When compared with KCS score model (AUC, 0.667; 95% CI, 0.625 to 0.710) with DNN model 2 in the internal validation set, DNN model 2 score (AUC, 0.730; 95% CI, 0.693 to 0.767) showed a better performance level than the KCS score model (p<0.001) (Fig. 4A). However, a comparison between the two models in the external validation set failed to show performance improvement with DNN model 2 (AUC, 0.765; 95% CI, 0.728 to 0.801) than KCS score model (AUC, 0.744; 95% CI, 0.707 to 0.780) (p=0.125) (Fig. 4B).

Table 2. Demographic and Clinical Data of the Study Participants

	Training set (n=49,235)	Tuning set (n=7,034)	Internal validation set (n=14,067)	External validation set (n=3,561)	p-value*
Age, yr	41.6±8.3	41.5±8.3	41.6±8.3	51.3±9.0	<0.001
Age group, yr					<0.001
<50	41,745 (84.8)	5,975 (84.9)	11,996 (85.3)	1,513 (42.5)	
50–69	7,275 (14.8)	1,015 (14.4)	1,987 (14.1)	1,959 (55.0)	
≥70	215 (0.4)	44 (0.6)	84 (0.6)	89 (2.5)	
Male sex	34,103 (69.3)	4,871 (69.3)	9,836 (69.9)	2,152 (60.4)	<0.001
Current or past smoker	13,992 (28.4)	1,964 (27.9)	4,018 (28.6)	1,698 (47.7)	<0.001
Family history of CRC	1,922 (3.9)	281 (4.0)	565 (4.0)	127 (3.6)	0.217
BMI, kg/m^2	23.8±3.1	23.8±3.1	23.8±3.1	23.8±3.1	0.227
BMI ≥25 kg/m^2	16,544 (33.6)	2,350 (33.4)	4,735 (33.7)	1,189 (33.4)	0.760
ACRN	693 (1.4)	86 (1.2)	181 (1.3)	169 (4.8)	<0.001
ACRN for age ≥50 yr	307 (4.1)	37 (3.5)	70 (3.4)	146 (7.1)	<0.001

Data are presented as mean±SD or number (%).
CRC, colorectal cancer; BMI, body mass index; ACRN, advanced colorectal neoplasia.
*Comparison between the internal and external validation sets.

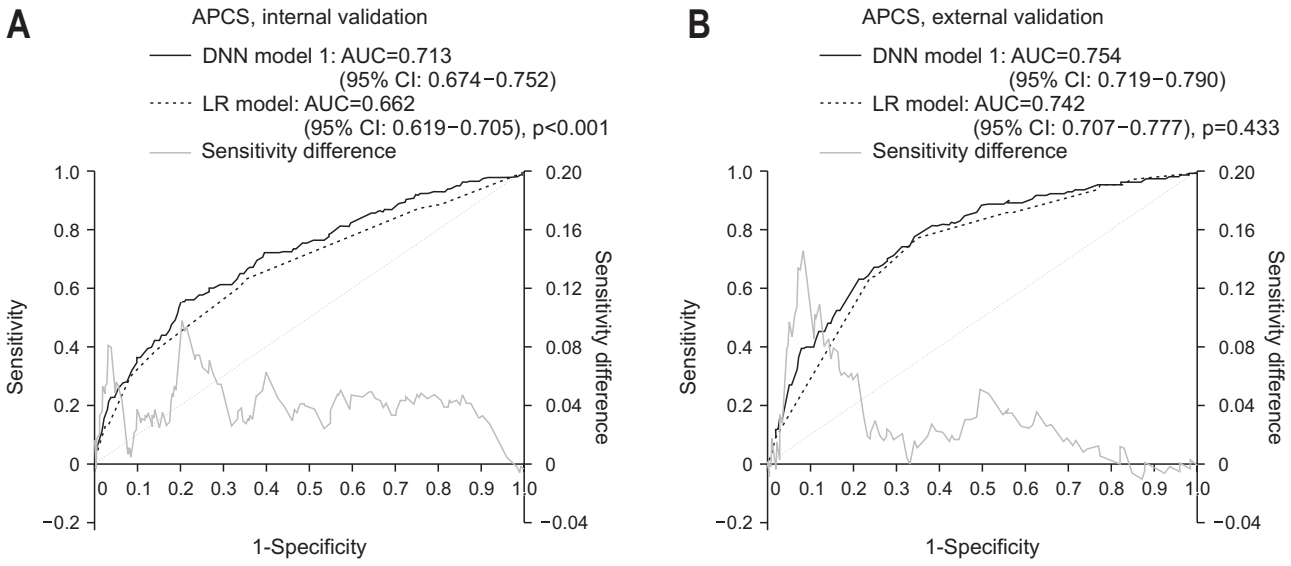

Fig. 3. Receiver operating characteristic curve and area under the curve of the prediction models for advanced colorectal neoplasia. Comparison between the Asian-Pacific Colorectal Screening (APCS) score model and deep neural network (DNN) model 1 in the internal validation set (A) and the external validation set (B).
AUC, area under the receiver operating characteristic curve; CI, confidence interval; LR, logistic regression.

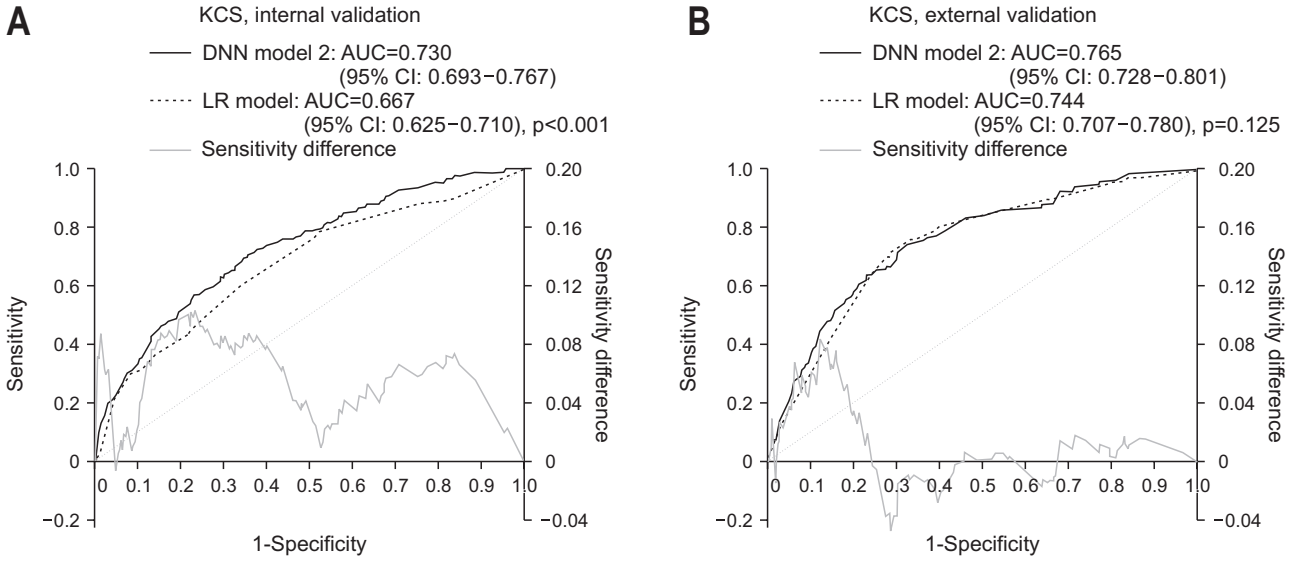

Fig. 4. Receiver operating characteristic curve and area under the curve of the prediction models for advanced colorectal neoplasia. Comparison between the Korean Colorectal Screening (KCS) score model and deep neural network (DNN) model 2 in the internal validation set (A) and the external validation set (B).
AUC, area under the receiver operating characteristic curve; CI, confidence interval; LR, logistic regression.

DISCUSSION

We expected better performance level to predict the risk of ACRN with DNN models than LR models because the interactions between the risk factors of ACRN would be complex and nonlinear to be reflected in the LR models. In this study, both DNN models 1 and 2 showed higher AUCs than the LR models for the APCS and KCS scores in the internal validation set as our expectation. However, both DNN models failed to show better performance than LR models in the external validation set. It may be explained by the limited use input variables (i.e., only 4–5 input variables) in this study. Therefore, a simple score model can predict the risk of ACRN as effectively as a DNN-based model, if the number of input variables is few.

Because of the suboptimal compliance of CRC screening, improved awareness of the personal risk of CRC may be helpful in increasing the screening rates.[31] Previously

used simple score models had an advantage of easy-to-use in daily clinical practice. However, they did not demonstrate good discriminative powers with maximum AUC or c-statistics ≤0.72.[5-9] In the LR methods, the inclusion of large numbers of covariates may lead to decreased model performance because of multiple collinearities or interactions.[32] In contrast, DNN-based models are promising because they may be able to capture the complex associations caused by the inclusion of large numbers of input parameters/nodes. As we have shown in this study, discriminative powers cannot be improved when only few input variables are used even with the DNN-based models. Therefore, this limitation should be considered to develop DNN-based algorithms for a risk prediction model.

Our findings should be considered with the limitations of DNN models. First, we adopted complete case analysis using DNN similar to the LR method. Therefore, our models used only four or five parameters. Even though DNN models have advantages that they can discover some structure in the training data and, consequently, incrementally modify data representation, resulting in superior accuracy of trained networks, this advantage of DNN models was blunted in our study as the training parameters are less than five. Second, our model did not specify when or the number of times the prediction of the risk of ACRN could be applied. Theoretically, these models could be applied at a specific age, such as 50 years. The age-specificity of these theoretical models should be evaluated in further studies before their application in CRC screening in real practice. Third, although the DNN models detected more ACRNs than the LR models did, the precise mechanisms of these models are not known. This black-box issue is important in clinical interpretations to identify why a specific individual was categorized as a high-risk of ACRN.[12] Fourth, there was a difference in the age distribution between cohort 1 of internal validation set and cohort 2 of external validation set. The internal validation set had more subjects under the age of 50 years than did the external validation set. This may be the reason for the DNN-based model not being better than the LR model in the external validation set. This difference in the age composition of the cohorts may limit the generalizability of the models to other populations.

In conclusion, simple score models for risk-prediction are as useful as DNN-based models with limited number of input variables. However, further studies on this topic are warranted to predict the risk of ACRN in CRC screening because DNN-based models are currently being developed and improved.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

ACKNOWLEDGEMENTS

This study was supported by a National Research Foundation (NRF) grant funded by the Korean government (NRF-2017R1A2B2009569).

AUTHOR CONTRIBUTIONS

Study concept and design: H.J.Y. Data acquisition: M.S.K. Data analysis and interpretation: C.W.C., S.K., K.S.A., S.K.P. Drafting of the manuscript; critical revision of the manuscript for important intellectual content; statistical analysis: J.K.M., H.J.Y. Obtained funding: H.J.Y. Administrative, technical, or material support; study supervision: J.M.C., D.I.P.

ORICD

Jun Ki Min https://orcid.org/0000-0003-0354-855X
Hyo-Joon Yang https://orcid.org/0000-0002-0265-672X
Min Seob Kwak https://orcid.org/0000-0002-8988-7423
Chang Woo Cho https://orcid.org/0000-0002-1430-8220
Sangsoo Kim https://orcid.org/0000-0001-9836-9823
Kwang-Sung Ahn https://orcid.org/0000-0002-2045-8348
Soo-Kyung Park https://orcid.org/0000-0001-8822-9632
Jae Myung Cha https://orcid.org/0000-0001-9403-230X
Dong Il Park https://orcid.org/0000-0003-2307-8575

REFERENCES

1. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 2015;136:E359-E386.
2. Shaukat A, Mongin SJ, Geisser MS, et al. Long-term mortality after screening for colorectal cancer. N Engl J Med 2013;369:1106-1114.
3. Nishihara R, Wu K, Lochhead P, et al. Long-term colorectal-cancer incidence and mortality after lower endoscopy. N Engl J Med 2013;369:1095-1105.
4. Liang PS, Wheat CL, Abhat A, et al. Adherence to competing strategies for colorectal cancer screening over 3 years. Am J

Gastroenterol 2016;111:105-114.
5. Imperiale TF, Monahan PO, Stump TE, Glowinski EA, Ransohoff DF. Derivation and validation of a scoring system to stratify risk for advanced colorectal neoplasia in asymptomatic adults: a cross-sectional study. Ann Intern Med 2015;163:339-346.
6. Yeoh KG, Ho KY, Chiu HM, et al. The Asia-Pacific Colorectal Screening score: a validated tool that stratifies risk for colorectal advanced neoplasia in asymptomatic Asian subjects. Gut 2011;60:1236-1241.
7. Kaminski MF, Polkowski M, Kraszewska E, Rupinski M, Butruk E, Regula J. A score to estimate the likelihood of detecting advanced colorectal neoplasia at colonoscopy. Gut 2014;63:1112-1119.
8. Lin OS, Kozarek RA, Schembre DB, et al. Risk stratification for colon neoplasia: screening strategies using colonoscopy and computerized tomographic colonography. Gastroenterology 2006;131:1011-1019.
9. Schroy PC 3rd, Wong JB, O'Brien MJ, Chen CA, Griffith JL. A risk prediction index for advanced colorectal neoplasia at screening colonoscopy. Am J Gastroenterol 2015;110:1062-1071.
10. Yang HJ, Choi S, Park SK, et al. Derivation and validation of a risk scoring model to predict advanced colorectal neoplasm in adults of all ages. J Gastroenterol Hepatol 2017;32:1328-1335.
11. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521: 436-444.
12. Min S, Lee B, Yoon S. Deep learning in bioinformatics. Brief Bioinform 2017;18:851-869.
13. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542:115-118.
14. Ting DSW, Cheung CY, Lim G, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. JAMA 2017;318:2211-2223.
15. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 2016;316:2402-2410.
16. Ehteshami Bejnordi B, Veta M, Johannes van Diest P, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA 2017;318:2199-2210.
17. Byrne MF, Chapados N, Soudan F, et al. Real-time differentiation of adenomatous and hyperplastic diminutive colorectal polyps during analysis of unaltered videos of standard colonoscopy using a deep learning model. Gut 2019;68:94-100.
18. Chen PJ, Lin MC, Lai MJ, Lin JC, Lu HH, Tseng VS. Accurate classification of diminutive colorectal polyps using computer-aided analysis. Gastroenterology 2018;154:568-575.
19. Kim DH, Cha JM, Shin HP, Joo KR, Lee JI, Park DI. Development and validation of a risk stratification-based screening model for predicting colorectal advanced neoplasia in Korea. J Clin Gastroenterol 2015;49:41-49.
20. Rhee EJ, Park SE, Chang Y, Ryu S, Lee WY. Baseline glycemic status and mortality in 241,499 Korean metropolitan subjects: a Kangbuk Samsung Health Study. Metabolism 2016;65:68-77.
21. Wen CP, David Cheng TY, et al. Are Asians at greater mortality risks for being overweight than Caucasians? Redefining obesity for Asians. Public Health Nutr 2009;12:497-506.
22. Tahmasebi P, Hezarkhani A. Application of a modular feedforward neural network for grade estimation. Nat Resour Res 2011;20:25-32.
23. Abadi M, Agarwal A, Barham P, et al. TensorFlow: large-scale machine learning on heterogeneous distributed systems. arXiv: 160304467v2 [Preprint]. 2016 [cited 2020 Mar 30]. Available from: https://arxiv.org/abs/1603.04467.
24. Aksoy S, Haralick RM. Feature normalization and likelihood-based similarity measures for image retrieval. Pattern Recognit Lett 2001;22:563-582.
25. Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv:14126980v9 [Preprint]. 2014 [cited 2020 Mar 30]. Available from: https://arxiv.org/abs/1412.6980.
26. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. J Mach Learn Res 2010;9:249-256.
27. Clevert DA, Unterthiner T, Hochreiter S. Fast and accurate deep network learning by exponential linear units (ELUs). arXiv:151107289v5 [Preprint]. 2015 [cited 2020 Mar 30]. Available from: https://arxiv.org/abs/1511.07289.
28. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 2014;15:1929-1958.
29. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988;44:837-845.
30. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 1983;148:839-843.
31. Sarfaty M, Wender R. How to increase colorectal cancer screening rates in practice. CA Cancer J Clin 2007;57:354-366.
32. Hosmer D, Lemeshow S. Applied logistic regression. 2nd ed. New York: Wiley, 2000.