ORIGINAL ARTICLE

Korean J Intern Med 2021;36:845-856
https://doi.org/10.3904/kjim.2020.020

KJIM

# Application of deep learning to predict advanced neoplasia using big clinical data in colorectal cancer screening of asymptomatic adults

Hyo-Joon Yang[1], Chang Woo Cho[2], Jongha Jang[2], Sang Soo Kim[2], Kwang-Sung Ahn[3], Soo-Kyung Park[1], and Dong Il Park[1]

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

**Received: January 19, 2020**
**Revised : February 28, 2020**
**Accepted: March 6, 2020**

**Correspondence to**
**Dong Il Park, M.D.**
Division of Gastroenterology, Department of Internal Medicine and Gastrointestinal Cancer Center, Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, 29 Saemunan-ro, Jongno-gu, Seoul 03181, Korea
Tel: +82-2-2001-8555
Fax: +82-2-2001-8360
E-mail: diksmc.park@samsung.com
https://orcid.org/0000-0003-2307-8575

**Background/Aims:** We aimed to develop a deep learning model for the prediction of the risk of advanced colorectal neoplasia (ACRN) in asymptomatic adults, based on which colorectal cancer screening could be customized.

**Methods:** We collected data on 26 clinical and laboratory parameters, including age, sex, smoking status, body mass index, complete blood count, blood chemistry, and tumor marker, from 70,336 first-time colonoscopy screening recipients. For reference, we used a logistic regression (LR) model with nine variables manually selected from the 26 variables. A deep neural network (DNN) model was developed using all 26 variables. The area under the receiver operating characteristic curve (AUC), sensitivity, and specificity of the models were compared in a randomly split validation group.

**Results:** In comparison with the LR model (AUC, 0.724; 95% confidence interval [CI], 0.684 to 0.765), the DNN model (AUC, 0.760; 95% CI, 0.724 to 0.795) demonstrated significantly improved performance with respect to the prediction of ACRN ($p < 0.001$). At a sensitivity of 90%, the specificity significantly increased with the application of the DNN model (41.0%) in comparison with the LR model (26.5%) ($p < 0.001$), indicating that the colonoscopy workload required to detect the same number of ACRNs could be reduced by 20%.

**Conclusions:** The application of DNN to big clinical data could significantly improve the prediction of ACRNs in comparison with the LR model, potentially realizing further customization by utilizing large quantities and various types of biomedical information.

**Keywords:** Colorectal neoplasms; Deep learning; Big data; Risk assessment; Mass screening

## INTRODUCTION

Colorectal cancer (CRC) is the third most common cancer worldwide and is more prevalent in developed countries [1]. Screening for CRC, which is recommended for the average risk population of 50 years and above, has been shown to reduce CRC-related mortality [2,3]. However, the effectiveness of this approach is affected by low adherence and inefficiency [4]. A large number of high-risk individuals have never been screened or have undergone non-invasive tests, resulting in wasted resources. Meanwhile, colonoscopy conducted on

pISSN 1226-3303
eISSN 2005-6648
http://www.kjim.org

low-risk individuals results in a low yield and leads to unnecessary complications. The customization of the screening based on the risk of CRC among the average risk population may improve the screening efficiency and adherence [5]. Several risk-prediction algorithms based on logistic regression (LR) models have been developed to identify individuals at high risk of advanced colorectal neoplasia (ACRN), for which colonoscopy may be most suitable [5-10]. However, these models demonstrated limited performance with low sensitivity and high false-positive rates, which may be due to the limited amount of information used in the model, limited performance of the LR method, or both.

Recently, deep learning has emerged as an alternative approach based on the accumulation of big data, advances in computational power, and improved algorithms [11]. It has outperformed previous machine learning techniques in various domains, including medicine [12]. Deep learning has shown expert-level accuracy in the diagnosis of skin cancer [13], diabetic retinopathy [14,15], lymph node metastasis of breast cancer [16], and colorectal adenoma during colonoscopy [17,18]. Clinical data are rapidly being obtained worldwide, and several laboratory parameters are reported to be associated with the risk of CRC [19]. However, a previous attempt that used an LR method to incorporate laboratory data into a risk model for identifying individuals at high risk of ACRN was not successful, with only minimal performance improvements being realized [10]. Deep learning may offer better prediction models for ACRN by utilizing big clinical data more efficiently than previous LR models.

Therefore, this study aimed to develop and validate a deep learning model for the prediction of the risk of ACRN in asymptomatic adults, and compare the developed model with an existing LR model with respect to CRC screening.

## METHODS

### Study population

This cross-sectional study was approved by the Institutional Review Board of Kangbuk Samsung Hospital (IRB No. 2017-07-024). The requirement for informed consent was exempted because only anonymized data were used.

We considered consecutive asymptomatic adults who underwent colonoscopy screening during health checkups at the Kangbuk Samsung Hospital Health Screening Center, Seoul, Korea, between January 2003 and December 2012. Exclusion criteria included previous colorectal examinations, such as barium enema, sigmoidoscopy, or colonoscopy, a history of CRC or other malignancies, a history of inflammatory bowel disease, a history of colorectal surgery, incomplete colonoscopy due to failed cecal intubation or inadequate bowel cleansing, and missing clinical data. The overall study population considered in the analysis was randomly split into development and validation groups in the ratio of 4:1.

### Dataset

From the health checkup results of the participants, 26 clinical and laboratory parameters were selected as input variables as well as colonoscopy data for the outcome variable.

As previously described [10,20], information on demographics, such as age, sex, and life style factors, was determined using standardized, self-administered questionnaires. For individuals with a family history of CRC, only first-degree relatives were considered, regardless of age. Trained nurses measured the physical parameters. According to the recommendation for Asians, a body mass index (BMI) ≥ 25 kg/m$^2$ was used to indicate obesity [21]. From the blood samples obtained after 10 hours of fasting, a range of laboratory parameters were measured.

In our screening program, some participants underwent a fecal immunochemical test (FIT) as well as colonoscopy [22]. A one-time stool sample was collected within 3 days before colonoscopy in a buffered sampling tube (Eiken Chemical Company, Tokyo, Japan) and sent to the laboratory on the day of the health examination. Fecal hemoglobin was quantified using an OC-SENSOR DIANA (Eiken Chemical Company) as ngHb/mL. A positive cut-off value of 100 ngHb/mL was considered equivalent to 20 μgHb/g feces.

Colonoscopy was conducted by 13 board-certified endoscopists using Evis Lucera CV-260 colonoscopes (Olympus Medical Systems, Tokyo, Japan). Colons were prepared with 4 L of polyethylene glycol solution. The endoscopists measured the sizes of all polyps, and then either performed a biopsy or removed them. Gastrointestinal pathologists evaluated histological specimens.

https://doi.org/10.3904/kjim.2020.020

ACRN was classified as colorectal carcinoma or advanced adenoma. Advanced adenoma was defined as any adenoma ≥ 1 cm in size, or one that has a villous component or high-grade dysplasia [10].

## Conventional machine learning methods

We first fitted an LR model to the development group for comparison (Supplementary Table 1) [10]. In a previous report, nine variables were manually selected from among 26 variables for this model as follows: age (< 50, 50 to 60, 60 to 70 vs. ≥ 70 years), sex, smoking status (none/past vs. current), family history of CRC, BMI (< 25 $kg/m^2$ vs. ≥ 25 $kg/m^2$), serum levels of fasting glucose (< 100 mg/L vs. ≥ 100 mg/L or diabetes), low-density lipoprotein-cholesterol (LDL-C; < 100 mg/L vs. ≥ 100 mg/L), and carcinoembryonic antigen (CEA; < 5 and 5 to 10 ng/mL vs. ≥ 10 ng/mL).

For *ad-hoc* analyses, we fitted another LR model that included all 26 variables. We further tested three conventional machine learning methods: support vector machine (SVM), random forest (RF), and extreme gradient boosting (XGBoost) for 26 variables [23,24].

## Development of deep neural networks

For deep learning, we used a feedforward neural network [25] as the deep neural network (DNN) structure, and Keras (version 2.2.4) [26] in Python (version 2.7.6.) as the deep learning framework. As illustrated in the Fig. 1A, we initially developed the DNN model using the same nine variables as that used in the LR model to determine whether deep learning could predict ACRN better than the LR method when the same information was provided. The main DNN model was developed using all 26 variables in the dataset as input nodes to clarify whether deep learning could overcome the limitations of the LR model, such as the compromise in prediction performance when a large number of covariates are considered. Moreover, all continuous variables were standardized for feature scaling [27]. For hyperparameter tuning, a 5-fold cross-validation was conducted (Fig. 1B) [28]. Consequently, the DNN with nine variables was set to have two hidden layers with 26 nodes for each layer, and the DNN with 26 variables was set to have two hidden layers with 10 nodes for each layer (Supplementary Table 2). Adam was used as an optimization algorithm with learning rate = 0.001, $\beta_1$ = 0.9, and $\beta_2$ = 0.999, as

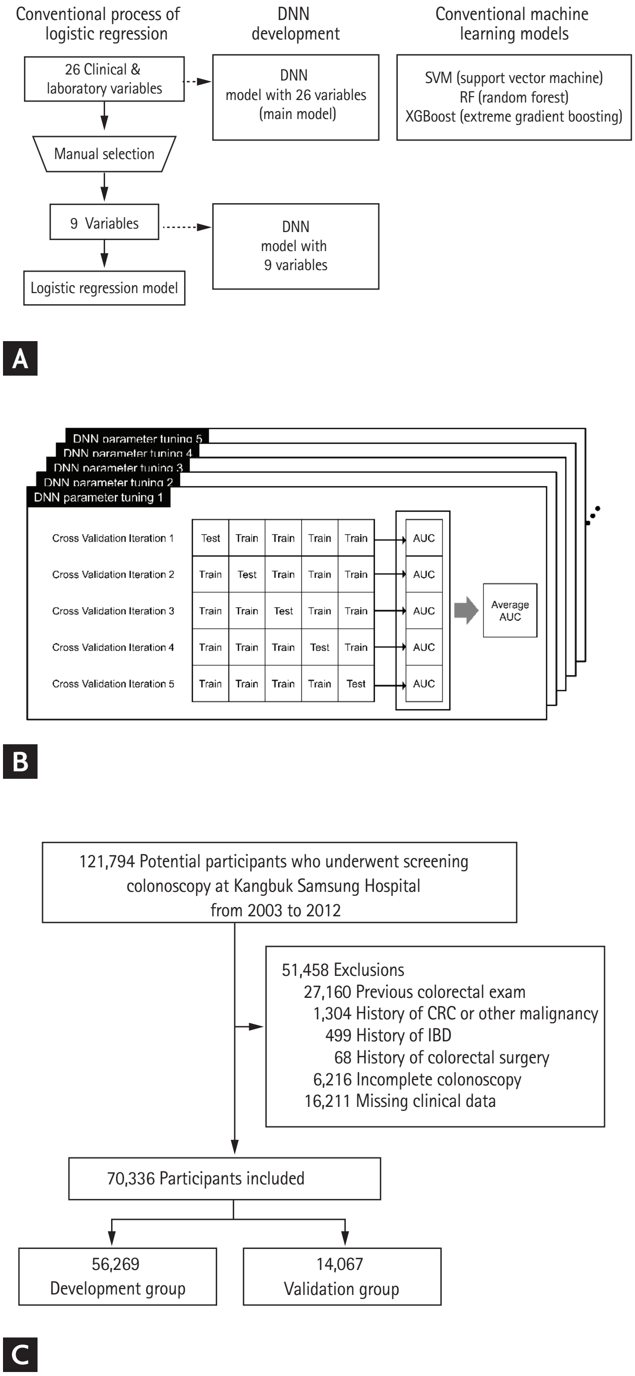
Conventional process of logistic regression
DNN development
Conventional machine learning models
26 Clinical & laboratory variables
DNN model with 26 variables (main model)
SVM (support vector machine)
RF (random forest)
XGBoost (extreme gradient boosting)
Manual selection
9 Variables
DNN model with 9 variables
Logistic regression model
A
DNN parameter tuning 5
DNN parameter tuning 4
DNN parameter tuning 3
DNN parameter tuning 2
DNN parameter tuning 1
Cross Validation Iteration 1 Test Train Train Train Train AUC
Cross Validation Iteration 2 Train Test Train Train Train AUC
Cross Validation Iteration 3 Train Train Test Train Train AUC
Cross Validation Iteration 4 Train Train Train Test Train AUC
Cross Validation Iteration 5 Train Train Train Train Test AUC
Average AUC
B
121,794 Potential participants who underwent screening colonoscopy at Kangbuk Samsung Hospital from 2003 to 2012
51,458 Exclusions
27,160 Previous colorectal exam
1,304 History of CRC or other malignancy
499 History of IBD
68 History of colorectal surgery
6,216 Incomplete colonoscopy
16,211 Missing clinical data
70,336 Participants included
56,269 Development group
14,067 Validation group
C

**Figure 1.** Deep learning model development process. (A) Conventional logistic regression process, deep neural network (DNN) model development, and conventional machine learning methods. (B) Cross-validation of DNN models. (C) Flow of the study population. SVM, support vector machine; RF, random forest; XGBoost, extreme gradient boosting; AUC, area under the receiver operating characteristic curve; CRC, colorectal cancer; IBD, inflammatory bowel disease.

proposed by Kingma and Ba [29]. The DNNs also applied the Xavier initializer [30] to initialize the weights of hidden units and the sigmoid activation function [31] in each layer. Binary cross-entropy was used to define a loss function [32]. We trained each model for 1,000 iterations using the dataset of the development group. The output value generated from the trained networks indicated the probability of each input case having ACRN, wherein the output ranged between 0 (low probability) and 1 (high probability).

### Statistical analysis

The primary analysis involved the comparison of the performance of the DNN model with that of the LR model for the prediction of ACRN in the validation group. The models were compared with respect to their area under the curve (AUC) of the receiver operating characteristic (ROC) curve using the DeLong test [33].

In a previous study on the LR model, the AUC was 0.68, and the prevalence of ACRN was 1.4% [10]. It was assumed that the detection of at least a 0.05 increment in the AUC in the DNN models would be clinically significant; therefore, it was estimated that at least 13,064 individuals would be required to detect this difference with 80% power, 5% significance level, and strong correlations (correlation coefficient, 0.7) between the models, both in the positive and negative cases [34].

From the perspective of CRC screening, model performances were also compared with respect to their sensitivity and specificity at three points with high sensitivity (80%, 90%, and 95%) on the ROC curve, which would be important for screening programs. At each point, the specificity and reduction in the number of colonoscopies needed to detect one ACRN (NNScope) for each method were estimated.

As *ad-hoc* analyses, we compared the LR and DNN models according to the number of included variables (nine vs. 26). SVM, RF, and XGBoost models as well as DNN were also compared with the LR model that included 26 variables as a reference. Furthermore, the performance of FIT and a combined FIT and clinical score, wherein colonoscopy is recommended for either individuals with positive FIT or a high-risk group in a clinical scoring model [10,35], were compared with the DNN model.

To address the black-box issue, which refers to the inability to learn how a DNN model predicts ACRNs [12], we compared the subjects that were predicted by both the LR and DNN models to have ACRN, those that were predicted only by the LR model, and those that were predicted only by the DNN model at the point of 90% sensitivity. Statistical analyses were performed using the R statistical programming environment, version 3.3.2 (R Development Core Team, Vienna, Austria; http://www.R-project.org). Furthermore, all *p* values were two-sided, and $p < 0.05$ was considered statistically significant.

## RESULTS

### Demographic and clinical characteristics of study population

During the study period, 121,794 individuals were screened. After excluding 51,458 individuals for reasons depicted in Fig. 1C, 70,336 individuals were included in the development group (n = 56,269) and validation group (n = 14,067). The mean age ± standard deviation (SD) of the overall study population was 41.6 ± 8.3 years, 69.4% (48,810/70,336) were male, and ACRN was detected in 1.4% (960/70,336) of the participants. The proportion of subjects aged 50 years or older was 15.1% (10,620/70,336), of which 3.9% (414/10,620) had ACRN. There were no significant differences between the demographics and clinical characteristics of the development and validation groups (Table 1). Although the differences in the serum glucose levels and high-sensitivity C-reactive protein (hsCRP) levels were statistically significant because of the large sample size, the actual difference had little clinical significance (development group vs. validation group: mean ± SD of glucose, 93.5 ± 14.6 mg/dL vs. 93.9 ± 15.5 mg/dL, $p < 0.007$; median [range] of hsCRP, 0.1 [0.0 to 0.1] mg/L vs. 0.1 [0.0 to 0.1] mg/L, $p = 0.038$).

### Performance of DNN model

The ROC curves of the LR and DNN models in the validation group are illustrated in Fig. 2. When compared with the LR model (AUC, 0.724; 95% confidence interval [CI], 0.684 to 0.765), the DNN model exhibited significantly improved performance (AUC, 0.760; 95% CI, 0.724 to 0.795; $p = 0.009$). The superiority of the DNN model over the LR model was prominent at the points with high sensitivity (≥ 80%) on the ROC curve. The

https://doi.org/10.3904/kjim.2020.020

**Table 1. Demographics and clinical characteristics of the study participants**

| Characteristic | Development group (n = 56,269) | Validation group (n = 14,067) | *p* value |
|---|---|---|---|
| Age, yr | 41.6 ± 8.3 | 41.6 ± 8.4 | 0.920 |
| Male sex | 39,063 (69.4) | 9,747 (69.3) | 0.761 |
| Current smoker | 15,930 (28.3) | 4,044 (28.8) | 0.303 |
| Alcohol consumption, time/wk | 2 (1–3) | 2 (1–3) | 0.660 |
| Regular exercise ≥ 4 times/wk | 30,666 (54.5) | 7,661 (54.5) | 0.935 |
| Family history of CRC | 2,202 (3.9) | 566 (4.0) | 0.547 |
| Hypertension | 9,386 (16.7) | 2,307 (16.4) | 0.424 |
| Diabetes | 2,827 (5.0) | 708 (5.0) | 0.965 |
| BMI, kg/m$^2$ | 23.8 ± 3.1 | 23.8 ± 3.1 | 0.360 |
| Waist circumference, cm | 83.2 ± 8.7 | 83.1 ± 8.6 | 0.393 |
| Systolic BP, mmHg | 113.3 ± 13.1 | 113.2 ± 13.0 | 0.453 |
| Diastolic BP, mmHg | 72.6 ± 9.6 | 72.5 ± 9.5 | 0.589 |
| Glucose, mg/dL | 93.5 ± 14.6 | 93.9 ± 15.5 | 0.007 |
| HbA1c, % | 5.7 ± 0.5 | 5.7 ± 0.5 | 0.981 |
| Total cholesterol, mg/dL | 199.8 ± 34.7 | 199.7 ± 34.7 | 0.638 |
| HDL-C, mg/dL | 55.1 ± 13.8 | 55.1 ± 13.8 | 0.601 |
| Triglyceride, mg/dL | 95 (67–141) | 96 (67–141) | 0.920 |
| LDL-C, mg/dL | 124.9 ± 32.0 | 124.9 ± 31.9 | 0.875 |
| Insulin, μU/mL | 4.5 (2.8–7.1) | 4.6 (2.8–7.1) | 0.946 |
| hsCRP, mg/L | 0.1 (0.0–0.1) | 0.1 (0.0–0.1) | 0.038 |
| WBC, × 10$^3$/mm$^3$ | 6.2 ± 1.7 | 6.2 ± 1.6 | 0.740 |
| RBC, × 10$^6$/mm$^3$ | 4.9 ± 0.4 | 4.9 ± 0.4 | 0.417 |
| Hemoglobin, g/dL | 14.9 ± 1.5 | 14.9 ± 1.5 | 0.642 |
| Hematocrit, % | 43.7 ± 4.0 | 43.7 ± 3.9 | 0.586 |
| Platelet, × 10$^3$/mm$^3$ | 248.2 ± 52.8 | 248.2 ± 52.5 | 0.990 |
| Ferritin, ng/mL | 139.6 (66.0–225.2) | 139.1 (64.8–221.9) | 0.142 |
| CEA, ng/mL | 1.4 (1.0–2.0) | 1.4 (1.0–2.0) | 0.683 |
| ACRN | 775 (1.4) | 185 (1.3) | 0.570 |
| ACRN for age ≥ 50 yr | 328/8,459 (3.9) | 86/2,161 (4.0) | 0.827 |

Values are presented as mean ± SD, number (%), or median (interquartile range).
CRC, colorectal cancer; BMI, body mass index; BP, blood pressure; HbA1c, hemoglobin A1c; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; hsCRP, high-sensitivity C-reactive protein; WBC, white blood cell; RBC, red blood cell; CEA, carcinoembryonic antigen; ACRN, advanced colorectal neoplasia.

performances of the prediction models with respect to CRC screening are presented in Table 2. At a sensitivity of 90%, with respect to detecting ACRNs, the specificity significantly increased with the application of the DNN model (41.0%) in comparison with the LR model (26.5%, $p < 0.001$). The DNN model could reduce the colonoscopy workload estimated by the NNScope that is required to detect the same number of ACRNs as the LR model by 19.9%. At other points with sensitivities of 80% and 95%, the DNN model demonstrated a slightly attenuated but still significant benefit over the LR model with a 13.8% and 8.4% reduced colonoscopy workload, respectively (both $p < 0.001$).

We further evaluated the prediction performance of the DNN model. First, we compared the LR and DNN models according to the number of variables (Fig. 3A). In comparison with the original LR model with nine variables (AUC, 0.724; 95% CI, 0.684 to 0.765), the LR model with 26 variables did not demonstrate any significant improvement in the performance (AUC, 0.734; 95% CI, 0.695 to 0.773). This value was lower than that for the DNN model with nine variables (AUC, 0.748; 95% CI, 0.711 to 0.784). Second, we compared SVM, RF, XGBoost, and DNN with the LR model with 26 variables for reference (Fig. 3B and Supplementary Table 3). In the validation group, only the DNN model exhibited a significantly better prediction performance than that of the LR model ($p = 0.036$). The SVM (AUC, 0.603; 95% CI, 0.556 to 0.649) and RF (AUC, 0.672; 95% CI, 0.632 to 0.712) exhibited a significantly lower prediction performance. The XGBoost exhibited a prediction performance as high (AUC, 0.760; 95% CI, 0.725 to 0.795) as that of the DNN model although it was not significantly better than that of the LR model ($p = 0.064$). Third, the performance of the DNN model was compared with that of FIT and the combined FIT and clinical score (Fig. 3C and Supplementary Table 4). The FIT results were available in 19.6% (2,751/14,067) of the validation group, and FIT was found to be positive in 2.9% (79/2,751). The sensitivity for ACRN was 27.3% and specificity was 97.4%. At the same sensitivity, the specificity of the DNN model was significantly lower at 90.5% ($p < 0.001$). The sensitivity of the combined FIT and clinical score was 42.4% and specificity was 90.7%. At the same sensitivity, the specificity of the DNN model was also significantly lower at 81.0% ($p < 0.001$).

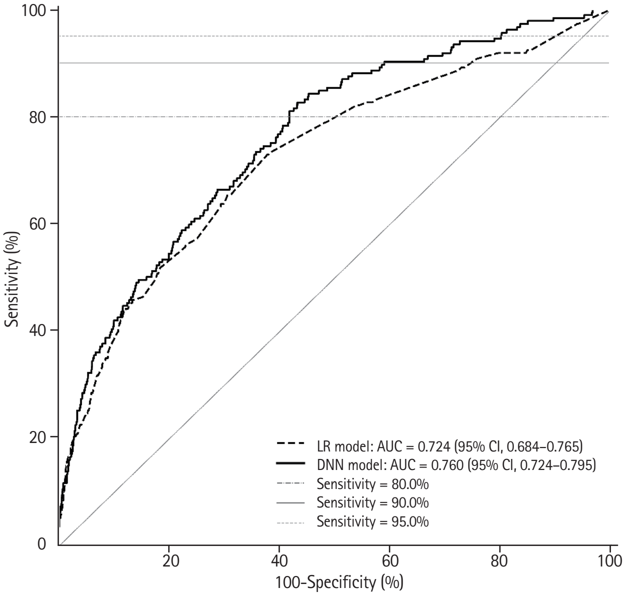

**Figure 2.** Receiver operating characteristic curve and area under the receiver operating characteristic curve (AUC) of the prediction models for advanced colorectal neoplasia. LR, logistic regression; DNN, deep neural network; CI, confidence interval.

**Table 2. Performance of DNN model at points of high sensitivity of detecting advanced neoplasia in colorectal cancer screening**

| Screening strategy | Sensitivity, % | Specificity, % | No. of colonoscopy | ACRNs detected, n | NNScope, n | Reduction of NNScope, % | *p* value (vs. LR) |
|---|---|---|---|---|---|---|---|
| Target sensitivity | 80 | | | | | | |
| LR | 78.9 | 51.7 | 6,855 | 146 | 47.0 | Reference | Reference |
| DNN | 79.5 | 58.2 | 5,948 | 147 | 40.5 | 13.8 | < 0.001 |
| Target sensitivity | 90 | | | | | | |
| LR | 89.2 | 26.5 | 10,364 | 165 | 62.8 | Reference | Reference |
| DNN | 89.7 | 41.0 | 8,356 | 166 | 50.3 | 19.9 | < 0.001 |
| Target sensitivity | 95 | | | | | | |
| LR | 92.4 | 14.5 | 12,041 | 171 | 70.4 | Reference | Reference |
| DNN | 94.6 | 19.9 | 11,293 | 175 | 64.5 | 8.4 | < 0.001 |

DNN, deep neural network; ACRN, advanced colorectal neoplasia; NNScope, number needed to colonoscope to detect one ACRN; LR, logistic regression.

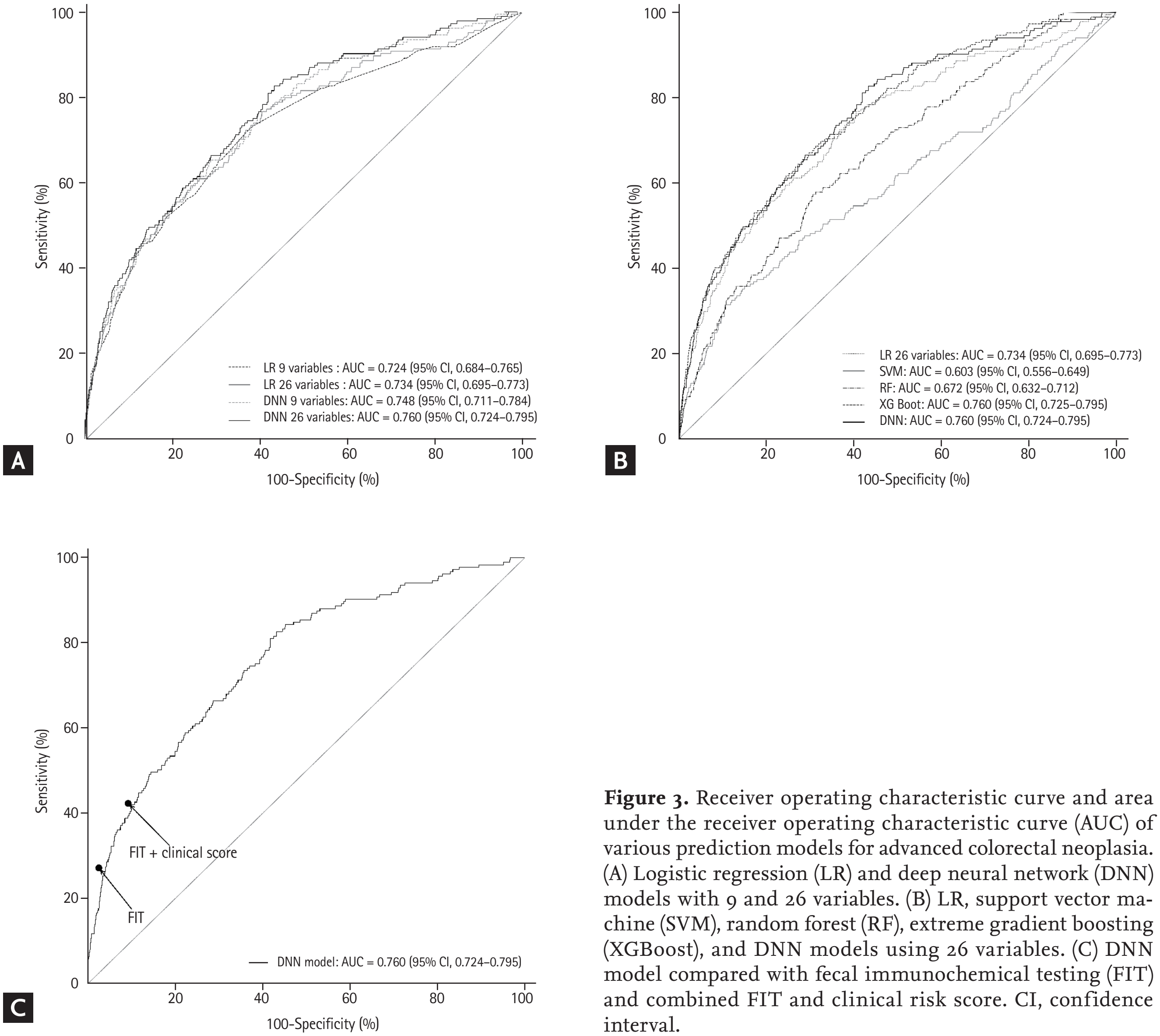

**Figure 3.** Receiver operating characteristic curve and area under the receiver operating characteristic curve (AUC) of various prediction models for advanced colorectal neoplasia. (A) Logistic regression (LR) and deep neural network (DNN) models with 9 and 26 variables. (B) LR, support vector machine (SVM), random forest (RF), extreme gradient boosting (XGBoost), and DNN models using 26 variables. (C) DNN model compared with fecal immunochemical testing (FIT) and combined FIT and clinical risk score. CI, confidence interval.

### Characteristics of the subjects with ACRNs detected by DNNs

At the target sensitivity of 90%, the actual number of subjects with ACRNs detected by the LR and DNN models were 165 and 166, respectively (Table 2). Most of them (n = 158) were detected using both the LR and DNN models. Meanwhile, seven subjects were detected using only the LR model, and eight using only the DNN model. To explore the additional features that could be captured by the DNN model, the three groups of subjects were compared based on their characteristics (Table 3). The participants with ACRNs who were predicted only by the DNN model were more likely to be women, had a lower BMI, higher serum levels of hsCRP, and lower levels of ferritin than those with ACRNs who were missed by the DNN model. When compared with those detected by both models, the subjects detected either by the LR or DNN model were younger and had lower serum levels of triglycerides.

## DISCUSSION

In this study, using a dataset of more than 70,000 subjects involving 26 clinical parameters, the DNN model exhibited better performance in the prediction of ACRN

**Table 3. Comparison of subjects with advanced colorectal neoplasia (ACRN) according to detection models**

| Characteristic | ACRNs detected both by LR and DNN (n = 157) | ACRNs detected only by LR model (n = 7) | ACRNs detected only by DNN model (n = 8) | *p* value |
|---|---|---|---|---|
| Age, yr | 51.5 ± 10.4 | 32.4 ± 4.8 | 43.4 ± 3.2 | 0.001 |
| Male sex | 138 (87.34) | 7 (100.0) | 2 (25.0) | < 0.001 |
| Current smoker | 65 (41.1) | 2 (28.6) | 3 (37.5) | 0.791 |
| Alcohol consumption, time/wk | 2 (1–3) | 1 (0–2) | 1.5 (1–2.5) | 0.214 |
| Regular exercise ≥ 4 times/wk | 88 (55.7) | 4 (57.4) | 3 (37.5) | 0.597 |
| Family history of CRC | 9 (5.7) | 0 (0) | 0 (0) | 0.637 |
| Hypertension | 35 (22.2) | 0 (0) | 2 (25.0) | 0.364 |
| Diabetes | 18 (11.4) | 0 (0) | 0 (0) | 0.385 |
| BMI, kg/m$^2$ | 24.6 ± 2.8 | 25.4 ± 5.1 | 22.1 ± 2.1 | 0.031 |
| Waist circumference, cm | 86.7 ± 7.4 | 86.4 ± 11.2 | 78.1 ± 4.8 | 0.100 |
| Systolic BP, mmHg | 113.7 ± 12.5 | 93.0 ± 5.2 | 108.5 ± 13.8 | 0.106 |
| Diastolic BP, mmHg | 72.6 ± 9.6 | 60.8 ± 5.0 | 70.1 ± 10.4 | 0.241 |
| Glucose, mg/dL | 99.2 ± 16.8 | 91.3 ± 11.4 | 86.8 ± 8.0 | 0.053 |
| HbA1c, % | 5.9 ± 0.6 | 5.5 ± 0.1 | 5.5 ± 0.2 | 0.065 |
| Total cholesterol, mg/dL | 209.8 ± 34.6 | 197.9 ± 28.8 | 172.8 ± 26.6 | 0.593 |
| HDL-C, mg/dL | 50.9 ± 12.3 | 50.6 ± 11.4 | 60.3 ± 16.4 | 0.510 |
| Triglyceride, mg/dL | 125 (93–191) | 100 (67–152) | 67 (52–92.5) | 0.004 |
| LDL-C, mg/dL | 134.7 ± 32.8 | 133.0 ± 27.9 | 100.0 ± 25.8 | 0.645 |
| Insulin, μU/mL | 5.5 (3.3–7.8) | 7.3 (2.5–10.6) | 3.5 (1.5–3.9) | 0.088 |
| hsCRP, mg/L | 0.1 (0.0–0.1) | 0.0 (0.0–0.1) | 0.1 (0.0–0.1) | 0.031 |
| WBC, × $10^3$/mm$^3$ | 6.9 ± 2.0 | 5.4 ± 0.8 | 7.4 ± 1.6 | 0.046 |
| RBC, × $10^6$/mm$^3$ | 4.9 ± 0.4 | 5.1 ± 0.2 | 4.4 ± 0.3 | 0.110 |
| Hemoglobin, g/dL | 15.4 ± 1.3 | 15.8 ± 0.6 | 13.9 ± 1.3 | 0.134 |
| Hematocrit, % | 45.2 ± 3.4 | 45.7 ± 2.3 | 41.3 ± 2.8 | 0.424 |
| Platelet, × $10^3$/mm$^3$ | 252.9 ± 58.1 | 223.9 ± 54.3 | 270.4 ± 38.8 | 0.433 |
| Ferritin, ng/mL | 154.2 (98.6–216.3) | 203.8 (140.9–326.8) | 81.9 (34.9–116.2) | 0.034 |
| CEA, ng/mL | 1.8 (1.2–2.5) | 1.2 (1.1–1.5) | 1.7 (1.1–2.8) | 0.295 |

Values are presented as mean ± SD, number (%), or median (interquartile range).
LR, logistic regression; DNN, deep neural network; BMI, body mass index; BP, blood pressure; HbA1c, hemoglobin A1c; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; hsCRP, high-sensitivity C-reactive protein; WBC, white blood cell; RBC, red blood cell; CEA, carcinoembryonic antigen.

in comparison with the conventional LR model. The value of AUC reached 0.76, which is higher than that of that of any other clinical prediction models or scores employed to predict ACRNs [5-10]. Importantly, this performance was achieved by the inclusion of 26 clinical and laboratory parameters, indicating the potential for the DNN to be expanded to include more data, even from other sources, such as transcriptomics and metabolomics information from blood, stool, tissue samples, or even imaging data. From the perspective of CRC screening, it was estimated that the use of our model could realize a reduction of 20% of the NNScope to detect the same number of ACRNs as the LR model.

The overall compliance of CRC screening remains suboptimal [4]. The improved awareness of personal risk of CRC may be helpful in increasing screening uptake [36]. However, previously reported clinical risk models did not demonstrate good discriminative power with maximum AUC or C-statistics ≤ 0.72 [5-9], and neither did the LR model with laboratory parameters, which was used as a reference in the current study [10]. In this study, the LR model with 26 variables exhibited a slight nonsignificant improvement in the AUC in comparison with the LR model with nine variables, from 0.72 to 0.73. In LR methods, the additional inclusion of a large number of covariates may not lead to a substantial improvement in the model performance because of multiple collinearities or interactions [37]. However, the application of DNN significantly improved the AUC from 0.72 of the LR model to 0.76 for the DNN model. This implies that the interactions between the risk factors for ACRNs may be too complex and nonlinear to be reflected by the LR models, whereas DNNs may be able to capture the complex associations caused by the inclusion of large numbers of input parameters/nodes [11]. The DNN has a multilayer architecture of input, hidden, and output layers. Each node of a hidden layer is computed as a function, which is usually nonlinear, of input nodes or previous hidden nodes that have their own weights. During each training example, the network is trained by updating the weights of the nodes through the backpropagation process. This multilayered structure of nonlinear functions and fine-tuned weights is capable of learning more complicated data structures. This is particularly important in modern times because of the substantial increase in the amount of biomedical information [12]. Furthermore, unlike the LR methods, a DNN can include various types of data as inputs, such as imaging data, fecal microbiome data, and electronic health record data [12]. In summary, we presented an enhanced performance DNN prediction model for ACRN, which may be able to improve adherence to CRC screening, indicating the possibility for further improvement by utilizing large quantities and various types of biomedical information.

It was estimated that colonoscopy resources are not sufficient, and tend to be overused in CRC screening [38,39]. Efficient screening can be achieved if ACRNs can be predicted with high specificity at a point of high sensitivity, which is associated with a lower colonoscopy workload being required for screening. In our study, the DNN model improved specificity by ≥ 80% of the sensitivity on the ROC curve. While ensuring that the number of ACRNs detected is not lesser than that detected by the LR model, the DNN model could reduce the NNScope by 20% in comparison with the LR model. Given the low marginal cost in the development of deep learning algorithms, our results imply that deep learning may promote a more efficient utilization of CRC screening resources without compromising health outcomes.

Our DNN model demonstrated significantly inferior specificity in comparison with FIT and the combined FIT and clinical score at the sensitivity points of 27% and 42%, respectively. However, FIT is limited by its low sensitivity unless the cut-off level is adjusted. In contrast, our model has the advantage of high specificity at the point of high sensitivity and the cut-off level can be chosen according to the available colonoscopy resources in individual societies.

Among the conventional machine learning methods, XGBoost exhibited a performance similar to that of the DNN model. XGBoost is an advanced implementation of the gradient boosting algorithm that is optimized for speed and performance [24]. Our results suggest that XGBoost could also potentially improve the prediction of ACRN in CRC screening. However, further study may be required to evaluate the role of XGBoost in the context of CRC screening, as this was not the focus of our study.

In this study, we observed several limitations of our deep learning model. First, although the DNN model detected more ACRNs than the LR model did, it is unknown as how the model actually functions. This black-box issue is important in clinical interpretations in terms of specifying why a specific individual was categorized as having a high risk of ACRN [12]. To address this issue, we reverse-engineered the DNN model, wherein the subjects with ACRNs who were detected both by the DNN and LR models, only by the LR model, and only by the DNN model were compared based on their clinical characteristics. The results demonstrated that the three groups differed substantially with respect to age, sex, BMI, triglycerides, white blood cell count, and ferritin. This implies that the DNN may result in a more accurate prediction by reducing the impact of conventional risk factors, in particular, sex and BMI. Second, we adopted a complete case analysis using deep learning similar to the LR method. Thus, our model considered 26 parameters, including serum glucose, LDL-C, and CEA, which are not directly applicable to other current CRC screening programs because not all these data are usually available in asymptomatic adults. However, it is not cost-effective to conduct laboratory analyses only for CRC screening. In the present study, we did not evaluate the potential for predicting ACRN or the degree of accuracy when only some of the parameters are given. Moreover, although we suggested the possibility of including various types of data, such as fecal microbiome data, such data were not available in our database. Thus, we could not demonstrate the feasibility of a model with 'omics' data. The answers to these questions are left to future research. Third, our model did not specify the time at which or the number of times that the prediction of ACRN could be applied. Theoretically, these models could be applied at a specific age, such as 40 or 50 years. Nevertheless, the age-specificity of these theoretical models need to be evaluated in further studies before their application to CRC screening in real practice.

In conclusion, the application of the DNN model to big clinical data significantly improved the prediction of ACRNs in comparison with the conventional LR model. This demonstrates the potential for realizing further performance improvements by utilizing large quantities and various types of biomedical information. This deep learning platform may accelerate the adoption of customized CRC screening based on the predicted risk of ACRN.

## KEY MESSAGE

1. A deep learning model demonstrated better performance than a conventional logistic regression model in the prediction of advanced colorectal neoplasia, by utilizing big clinical data more efficiently.
2. With the application of the deep learning model, the colonoscopy workload required to detect the same number of advanced colorectal neoplasia could be reduced by 20%.
3. Deep learning offers the potential for further improvements by utilizing large quantities and various types of biomedical information.

### Conflict of interest

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

### Acknowledgments

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

## 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.

https://doi.org/10.3904/kjim.2020.020

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. Ahmed RL, Schmitz KH, Anderson KE, Rosamond WD, Folsom AR. The metabolic syndrome and risk of incident colorectal cancer. Cancer 2006;107:28-36.
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, Tsai SP, 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. Jung YS, Park CH, Kim NH, Park JH, Park DI, Sohn CI. Identifying the optimal strategy for screening of advanced colorectal neoplasia. J Gastroenterol Hepatol 2017;32:1003-1010.
23. Futoma J, Morris J, Lucas J. A comparison of models for predicting early hospital readmissions. J Biomed Inform 2015;56:229-238.
24. Ogunleye AA, Qing-Guo W. XGBoost model for chronic kidney disease diagnosis. IEEE/ACM Trans Comput Biol Bioinform 2020;17:2131-2140.
25. Tahmasebi P, Hezarkhani A. Application of a modular feedforward neural network for grade estimation. Nat Resour Res 2011;20:25-32.
26. Keras. Keras: the python deep learning library [Internet]. c2017 [cited 2020 Jul 16]. Available from: https://keras.io/.
27. Aksoy S, Haralick RM. Feature normalization and likelihood-based similarity measures for image retrieval. Pattern Recognit Lett 2001;22:563-582.
28. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: IJCAI. Proceedings of the 14th International Joint Conference on Artificial Intelligence, Vol. 2; 1995 Aug 20-25; Montreal, QC. Montreal (QC): IJCAI, 1995: 1137-1143.
29. Kingma DP, Ba JL. Adam: a method for stochastic optimization. arXiv 2014. https://arxiv.org/abs/1412.6980.
30. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. Proc Mach Learn Res 2010;9:249-256.
31. Gibbs MN, MacKay DC. Variational Gaussian process classifiers. IEEE Trans Neural Netw 2000;11:1458-1464.
32. Alon G, Kroese DP, Raviv T, Rubinstein RY. Application of the cross-entropy method to the buffer allocation problem in a simulation-based environment. Ann Oper Res 2005;134:137-151.
33. 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.
34. 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.
35. Chiu HM, Ching JY, Wu KC, et al. A risk-scoring system combined with a fecal immunochemical test is effective in screening high-risk subjects for early colonoscopy to detect advanced colorectal neoplasms. Gastroenterology 2016;150:617-625.
36. Sarfaty M, Wender R. How to increase colorectal cancer screening rates in practice. CA Cancer J Clin 2007;57:354-

366.

37. Hosmer Jr DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression. 3rd ed. Hoboken (NJ): Wiley & Sons, 2013.

38. Levin TR. Colonoscopy capacity: can we build it? Will they come? Gastroenterology 2004;127:1841-1844.

39. Kruse GR, Khan SM, Zaslavsky AM, Ayanian JZ, Sequist TD. Overuse of colonoscopy for colorectal cancer screening and surveillance. J Gen Intern Med 2015;30:277-283.

**Supplementary Table 1. Multiple logistic regression model fitted in the development group**

| Covariate | β coefficient | Odds ratio | 95% CI | *p* value |
|---|---|---|---|---|
| Age group, yr | | | | < 0.001 |
| < 50 | 1 | 1 | | |
| ≥ 50, < 60 | 1.242 | 3.46 | 2.91–4.12 | |
| ≥ 60, < 70 | 1.864 | 6.45 | 5.17–8.06 | |
| ≥ 70 | 2.298 | 9.95 | 6.33–15.6 | |
| Male sex | 0.451 | 1.57 | 1.32–1.87 | 0.001 |
| Current smoker | 0.279 | 1.32 | 1.14–1.54 | 0.001 |
| Family history of CRC | 0.050 | 1.05 | 0.76–1.46 | 0.765 |
| BMI (≥ 25 kg/m$^2$) | 0.266 | 1.30 | 1.12–1.51 | 0.001 |
| Glucose (≥ 100 mg/dL or diabetes) | 0.179 | 1.20 | 1.02–1.40 | 0.025 |
| LDL-C (≥ 100 mg/dL) | 0.177 | 1.19 | 0.99–1.45 | 0.069 |
| CEA group, ng/mL | | | | 0.001 |
| < 5 | 1 | 1 | | |
| ≥ 5, < 10 | 0.867 | 2.38 | 1.59–3.56 | |
| ≥ 10 | 2.174 | 8.79 | 3.47–22.31 | |
| Constant | –5.405 | 0.00 | 0.00–0.01 | < 0.001 |

CI, confidence interval; CRC, colorectal cancer; BMI, body mass index; LDL-C, low-density lipoprotein cholesterol; CEA, carcinoembryonic antigen.

**Supplementary Table 2. Comparison of performances of deep neural network models with different values of hyperparameters**

| Model no. | Hidden layer, n | Node at layer #1, n | Node at layer #2, n | Node at layer #3, n | Node at layer #4, n | AUC in CV set 1 | AUC in CV set 2 | AUC in CV set 3 | AUC in CV set 4 | AUC in CV set 5 | Average AUC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Deep neural network with nine variables | | | | | | | | | | | |
| 1 | 2 | 26 | 26 | - | - | 0.753 | 0.730 | 0.733 | 0.718 | 0.731 | 0.7330 |
| 2 | 2 | 25 | 20 | - | - | 0.753 | 0.730 | 0.732 | 0.720 | 0.730 | 0.7330 |
| 3 | 2 | 25 | 25 | - | - | 0.753 | 0.729 | 0.733 | 0.720 | 0.730 | 0.7330 |
| 4 | 3 | 26 | 26 | 26 | - | 0.750 | 0.732 | 0.734 | 0.718 | 0.728 | 0.7324 |
| 5 | 3 | 24 | 24 | 20 | - | 0.752 | 0.732 | 0.734 | 0.720 | 0.724 | 0.7324 |
| 6 | 3 | 25 | 25 | 25 | - | 0.752 | 0.732 | 0.733 | 0.720 | 0.724 | 0.7322 |
| 7 | 3 | 25 | 20 | 10 | - | 0.752 | 0.733 | 0.733 | 0.720 | 0.723 | 0.7322 |
| 8 | 3 | 20 | 10 | 5 | - | 0.751 | 0.729 | 0.734 | 0.720 | 0.727 | 0.7322 |
| 9 | 2 | 28 | 28 | - | - | 0.751 | 0.730 | 0.733 | 0.717 | 0.729 | 0.7320 |
| 10 | 3 | 20 | 12 | 3 | - | 0.750 | 0.732 | 0.732 | 0.719 | 0.727 | 0.7320 |
| 11 | 2 | 30 | 30 | - | - | 0.751 | 0.730 | 0.731 | 0.718 | 0.729 | 0.7318 |
| 12 | 4 | 26 | 26 | 26 | 26 | 0.751 | 0.731 | 0.734 | 0.718 | 0.723 | 0.7314 |
| 13 | 2 | 18 | 8 | - | - | 0.751 | 0.729 | 0.731 | 0.715 | 0.731 | 0.7314 |
| 14 | 3 | 20 | 15 | 5 | - | 0.750 | 0.730 | 0.731 | 0.720 | 0.725 | 0.7312 |
| 15 | 3 | 20 | 11 | 4 | - | 0.751 | 0.730 | 0.731 | 0.720 | 0.722 | 0.7308 |
| 16 | 3 | 20 | 13 | 2 | - | 0.752 | 0.729 | 0.723 | 0.720 | 0.730 | 0.7308 |
| 17 | 3 | 30 | 20 | 10 | - | 0.750 | 0.732 | 0.732 | 0.719 | 0.720 | 0.7306 |
| 18 | 2 | 16 | 8 | - | - | 0.750 | 0.729 | 0.726 | 0.717 | 0.730 | 0.7304 |
| 19 | 2 | 10 | 2 | - | - | 0.752 | 0.728 | 0.724 | 0.715 | 0.732 | 0.7302 |
| 20 | 2 | 12 | 4 | - | - | 0.750 | 0.728 | 0.726 | 0.716 | 0.731 | 0.7302 |
| 21 | 4 | 25 | 25 | 25 | 25 | 0.748 | 0.730 | 0.734 | 0.718 | 0.720 | 0.7300 |
| 22 | 2 | 14 | 6 | - | - | 0.751 | 0.729 | 0.727 | 0.715 | 0.725 | 0.7294 |
| 23 | 2 | 20 | 10 | - | - | 0.737 | 0.734 | 0.714 | 0.725 | 0.722 | 0.7264 |
| Deep neural network with 26 variables | | | | | | | | | | | |
| 1 | 2 | 10 | 10 | - | - | 0.742 | 0.740 | 0.719 | 0.730 | 0.727 | 0.7316 |
| 2 | 3 | 20 | 20 | 20 | - | 0.745 | 0.737 | 0.718 | 0.723 | 0.722 | 0.729 |
| 3 | 2 | 15 | 15 | - | - | 0.746 | 0.738 | 0.720 | 0.724 | 0.716 | 0.7288 |
| 4 | 2 | 20 | 20 | - | - | 0.745 | 0.738 | 0.723 | 0.725 | 0.713 | 0.7288 |
| 5 | 2 | 25 | 25 | - | - | 0.745 | 0.739 | 0.718 | 0.722 | 0.720 | 0.7288 |
| 6 | 3 | 25 | 25 | 25 | - | 0.749 | 0.733 | 0.718 | 0.726 | 0.716 | 0.7284 |
| 7 | 3 | 22 | 22 | 22 | - | 0.743 | 0.736 | 0.722 | 0.722 | 0.719 | 0.7284 |
| 8 | 3 | 35 | 35 | 35 | - | 0.748 | 0.736 | 0.718 | 0.722 | 0.715 | 0.7278 |
| 9 | 4 | 50 | 40 | 30 | 10 | 0.740 | 0.734 | 0.718 | 0.721 | 0.722 | 0.7270 |
| 10 | 2 | 40 | 40 | - | - | 0.742 | 0.739 | 0.723 | 0.723 | 0.708 | 0.7270 |
| 11 | 3 | 45 | 45 | 45 | - | 0.745 | 0.733 | 0.720 | 0.721 | 0.715 | 0.7268 |
| 12 | 4 | 30 | 30 | 30 | 30 | 0.738 | 0.734 | 0.714 | 0.725 | 0.722 | 0.7266 |
| 13 | 4 | 30 | 30 | 30 | 30 | 0.738 | 0.734 | 0.714 | 0.725 | 0.722 | 0.7266 |
| 14 | 3 | 55 | 55 | 55 | - | 0.741 | 0.736 | 0.718 | 0.720 | 0.717 | 0.7264 |
| 15 | 4 | 25 | 25 | 25 | 25 | 0.739 | 0.733 | 0.715 | 0.725 | 0.720 | 0.7264 |

**Supplementary Table 2.Continued**

| Model no. | Hidden layer, n | Node at layer #1, n | Node at layer #2, n | Node at layer #3, n | Node at layer #4, n | AUC in CV set 1 | AUC in CV set 2 | AUC in CV set 3 | AUC in CV set 4 | AUC in CV set 5 | Average AUC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 16 | 3 | 10 | 10 | 10 | - | 0.739 | 0.742 | 0.713 | 0.717 | 0.720 | 0.7262 |
| 17 | 4 | 20 | 20 | 20 | 20 | 0.739 | 0.735 | 0.716 | 0.723 | 0.718 | 0.7262 |
| 18 | 3 | 65 | 65 | 65 | - | 0.741 | 0.736 | 0.716 | 0.720 | 0.717 | 0.7260 |
| 19 | 4 | 40 | 40 | 40 | 40 | 0.736 | 0.734 | 0.714 | 0.728 | 0.717 | 0.7258 |
| 20 | 4 | 45 | 45 | 45 | 45 | 0.74 | 0.732 | 0.716 | 0.719 | 0.719 | 0.7252 |
| 21 | 4 | 35 | 35 | 35 | 35 | 0.739 | 0.732 | 0.716 | 0.722 | 0.716 | 0.7250 |
| 22 | 4 | 60 | 40 | 20 | 5 | 0.739 | 0.731 | 0.714 | 0.720 | 0.719 | 0.7246 |
| 23 | 4 | 40 | 30 | 20 | 10 | 0.740 | 0.731 | 0.714 | 0.722 | 0.715 | 0.7244 |

All deep neural networks presented in the table used the sigmoid functions, the Xavier initializer, and the Adam optimizer with 1,000 epochs of the same dataset of the development group for each model.

AUC, area under the receiver operating characteristic curve; CV, cross-validation.

**Supplementary Table 3. Performance of conventional machine learning methods as well as DNN model compared with LR with 26 variables in the prediction of advanced neoplasia**

| Prediction model | AUC (95% CI) | *p* value |
|---|---|---|
| LR with 26 variables | 0.734 (0.695–0.773) | Reference |
| SVM | 0.603 (0.556–0.649) | < 0.001 |
| RF | 0.672 (0.632–0.712) | 0.001 |
| XGBoost | 0.760 (0.725–0.795) | 0.064 |
| DNN | 0.760 (0.724–0.795) | 0.036 |

DNN, deep neural network; LR, logistic regression; AUC, area under the receiver operating characteristic curve; CI, confidence interval; SVM, support vector machine; RF, random forest; XGBoost, extreme gradient boosting.

https://doi.org/10.3904/kjim.2020.020

**Supplementary Table 4. Performance of DNN model compared with FIT at points of low sensitivity of detecting advanced neoplasia**

| Screening strategy | Sensitivity, % | Specificity, % | No. of colonoscopy | ACRNs detected, n | NNScope, n | *p* value |
|---|---|---|---|---|---|---|
| Target sensitivity | 27.3 | | | | | |
| DNN | 27.3 | 90.5 | 266 | 9 | 29.6 | Reference |
| FIT | 27.3 | 97.4 | 79 | 9 | 8.8 | < 0.001 |
| Target sensitivity | 42.4 | | | | | |
| DNN | 39.4 | 81.0 | 529 | 13 | 40.7 | Reference |
| Combined FIT and clinical score | 42.4 | 90.7 | 267 | 14 | 19.1 | < 0.001 |

DNN, deep neural network; FIT, fecal immunochemical test; ACRN, advanced colorectal neoplasia; NNScope, number needed to colonoscope to detect one advanced colorectal neoplasia.