Usefulness of Deep Learning Analysis for the Diagnosis of Malignancy in Intraductal Papillary Mucinous Neoplasms of the Pancreas

Takamichi Kuwahara, MD, PhD[1], Kazuo Hara, MD, PhD[1], Nobumasa Mizuno, MD, PhD[1], Nozomi Okuno, MD[1], Shimpei Matsumoto, MD[1], Masahiro Obata, MD[1], Yusuke Kurita, MD[1], Hiroki Koda, MD[1], Kazuhiro Toriyama, MD[1], Sachiyo Onishi, MD[2], Makoto Ishihara, MD, PhD[2], Tsutomu Tanaka, MD, PhD[2], Masahiro Tajika, MD, PhD[2] and Yasumasa Niwa, MD, PhD[2]

OBJECTIVES: **Intraductal papillary mucinous neoplasms (IPMNs) are precursor lesions of pancreatic adenocarcinoma. Artificial intelligence (AI) is a mathematical concept whose implementation automates learning and recognizing data patterns. The aim of this study was to investigate whether AI *via* deep learning algorithms using endoscopic ultrasonography (EUS) images of IPMNs could predict malignancy.**

METHODS: **This retrospective study involved the analysis of patients who underwent EUS before pancreatectomy and had pathologically confirmed IPMNs in a single cancer center. In total, 3,970 still images were collected and fed as input into the deep learning algorithm. AI value and AI malignant probability were calculated.**

RESULTS: **The mean AI value of malignant IPMNs was significantly greater than benign IPMNs (0.808 vs 0.104, $P < 0.001$). The area under the receiver operating characteristic curve for the ability to diagnose malignancies of IPMNs *via* AI malignant probability was 0.98 ($P < 0.001$). The sensitivity, specificity, and accuracy of AI malignant probability were 95.7%, 92.6%, and 94.0%, respectively; its accuracy was higher than human diagnosis (56.0%) and the mural nodule (68.0%). Multivariate logistic regression analysis showed AI malignant probability to be the only independent factor for IPMN-associated malignancy (odds ratio: 295.16, 95% confidence interval: 14.13–6,165.75, $P < 0.001$).**

DISCUSSION: **AI *via* deep learning algorithm may be a more accurate and objective method to diagnose malignancies of IPMNs in comparison to human diagnosis and conventional EUS features.**

Clinical and Translational Gastroenterology 2019;10:e-00045. https://doi.org/10.14309/ctg.0000000000000045

INTRODUCTION

Intraductal papillary mucinous neoplasms (IPMNs) are precursor lesions of pancreatic adenocarcinoma (1). Once IPMNs progress to invasive cancer, the prognosis may be as poor as conventional pancreatic ductal adenocarcinoma. Resection of IPMNs, particularly in the stage of high-grade dysplasia, is presumed to provide a survival benefit (2). Endoscopic ultrasonography (EUS), which could evaluate the pancreas with high accuracy, is used to assess the malignancy of IPMNs. The international consensus guidelines for the management of IPMNs were proposed in 2012 and revised in 2017 (3,4). In the guidelines, high-risk stigmata (HRS) that were highly suspected as malignant and worrisome features (WFs) that were suspected as malignant were defined. Three criteria in HRS and 8 in WF were also developed. The guidelines recommended the use of HRS and WF to determine the medical treatment of IPMNs. The diagnostic accuracy in detecting the malignancy of IPMNs was evaluated using the guidelines, but it was not highly sufficient (5–7). In the European guidelines, conservative management and absolute and relative indications for surgery in IPMN cases were defined using potential prognostic factors (8). Several predictive techniques, such as logistic regression analysis, nomogram, cyst fluid analysis, and gene analysis, were used to diagnose the malignancy of IPMNs more precisely. However, these techniques did not show highly satisfactory results (70%–80%) (9–13).

Artificial intelligence (AI) is a mathematical predicting technique that automates learning and recognizing data patterns. Deep learning is an AI algorithm and advanced type of machine learning

[1]Department of Gastroenterology, Aichi Cancer Center Hospital, Nagoya, Japan; [2]Department of Endoscopy, Aichi Cancer Center Hospital, Nagoya, Japan.
Correspondence: Takamichi Kuwahara, MD, PhD. E-mail: kuwa_tak@aichi-cc.jp.
Received September 23, 2018; accepted February 28, 2019; published online May 7, 2019

method that uses neural networks (14). Deep learning provides a high-performance prediction. It is frequently used for AI algorithms and has been applied for medical diagnosis (15–18).

The aim of this study was to investigate whether preoperative AI *via* the deep learning algorithm using EUS images of IPMN could predict the diagnosis of malignancy and compare the diagnostic ability of IPMN malignancy *via* AI with that of human preoperative diagnosis, conventional predictive techniques, conventional EUS features, and other prognostic factors that were reported in the guidelines.

METHODS

Patients

From June 1995 to September 2017, a retrospective study was performed on 206 patients who underwent EUS before pancreatic resection and had pathologically confirmed IPMN after the surgery. The patients whose EUS images of IPMN were recorded in a digital format were included in this study. The following features were evaluated: age at the time of the operation, sex, tumor location, clinical symptoms (including history of pancreatitis), preoperative laboratory values (serum amylase [AMY], carcinoembryonic antigen [CEA], and carbohydrate antigen 19-9 [CA19-9] levels), imaging findings (mural nodule size, main pancreatic duct [MPD] diameter, and cyst size), and pathological findings. EUS was an essential preoperative assessment for all patients. It was used to determine the mural nodule size, MPD diameter, cyst size, and growth rate. All mural nodules were confirmed using contrast-enhanced EUS and/or computed tomography. Human preoperative diagnosis was defined as the preoperative diagnosis that doctors judged comprehensively using clinical information, laboratory values, and image findings. Pathological diagnosis of IPMN was classified as low-grade dysplasia, intermediate-grade dysplasia, high-grade dysplasia, and invasive carcinoma. Invasive carcinoma was defined as a histological transition that was clearly present between the IPMN and pancreatic ductal adenocarcinoma. All regions were categorized as benign (low- and intermediate-grade dysplasia) and malignant (high-grade dysplasia and invasive carcinoma) on the basis of the pathological diagnosis after resection. To compare the diagnostic performance of AI, human preoperative diagnosis and conventional logistic regression analysis using conventional EUS features and other prognostic factors that were reported in the guidelines were evaluated (8,9).

This study was approved by the Institutional Review Board of the Aichi Cancer Center (No. 2016-1-367, date: April 14, 2017) and performed in accordance with the Declaration of Helsinki (19).

EUS procedure

In all patients, EUS was performed using SSD-5500 or Prosound SSD α-10 (Hitachi Aloka Medical, Tokyo, Japan) and EU-ME2 (Olympus Corporation, Tokyo, Japan) ultrasound system with GF-UC30P, GF-UC240P-AL5, GF-UCT260, or GF-UCT240 curved linear echoendoscope (Olympus Corporation). All patients underwent EUS, and a video clip of EUS images was recorded. From these images, all images of the IPMNs were stored as digital still images (JPEG format).

Deep learning algorithm

TensorFlow version 1.8 (Google LCC, Mountain View) was used for the deep learning algorithm. Deep learning is the process of training a neural network (a large mathematical function with millions of parameters) to perform a task (14). A neural network is a machine learning technique that outputs the result mathematically after inputting numeric values or image information. A neural network consists of an input layer, a hidden layer, and an output layer. All layers are connected in series or parallel. Input data were converted to the output data by applying a weight to the input data, adding the bias, and passing to the activation function at each layer (Figure 1a). A neural network with multiple hidden layers is called deep learning. During the training process of deep learning, labeled information was put into the algorithm, and the output values were then calculated. In this study, EUS images of IPMN that were diagnosed pathologically were used for input information. All EUS images were trimmed to the same size square. After that, EUS images were converted into levels on a gray scale (0–255) in each pixel. As a result, all EUS images were converted to the mathematical information and were put into the algorithm. The parameters of the algorithm (biases and weights) were adjusted mathematically to decrease the error between the real results and the output values. This process is called "training," which uses the optimization algorithm and is repeated many times on each image in the training set (18). After training, the deep learning algorithm is completed, and test data are evaluated by this algorithm.

The convolutional neural network (CNN) was the specific neural network architecture used in this work. The CNN has proven to be an effective model for a variety of visual tasks (20). In the CNN, each pixel value of the input images was converted to the feature maps by multiplying the filter weights and sliding the filter over the input images (Figure 1b). Based on the CNN technique, several high-performance algorithms, such as AlexNet (20), GoogleNet (21), VGG16 (22), and ResNet (23), were generated. These algorithms are composed of several CNN layers, other layers such as max pooling, global average pooling, or fully connected layers (5–100 layers), and several activation functions. ResNet is composed of residual blocks in which there are shortcut connections between the CNN layers (Figure 1c). In this study, the original deep learning algorithm that was based on the ResNet50 algorithm was used (Figure 2) (23). The data were labeled according to the manner in which IPMN was defined (malignant, 1; benign, 0), according to the pathological results. The EUS images of IPMN were input and then processed by CNN, max pooling, and global average pooling layers. Swish activation functions (24) were used for the hidden layers and softmax function for the output layers. To speed up the training, batch normalization (25) was used. To prevent overfitting, stochastic depth (26), early stopping (27), data augmentation (28), random cropping (20), and random erasing (29) were used. The optimization algorithm used to train the network weights was a momentum stochastic gradient descent estimation implementation (23). After training, the output value of deep learning was calculated as the predictive value of malignant probability using AI (AI value: continuous variables from 0 to 1) in the test set. When the AI value became close to 1, malignant probability became increased. In this study, each EUS image was put into the deep learning algorithm, and the AI values were the output. To compare the diagnostic ability of AI with other conventional EUS features, AI malignant probability, which was defined as the mean AI value of all images in each patient, was calculated. In this study, a 10-fold cross-validation (training/test set ratio: 90%/10% × 10) was used to verify the validity of this algorithm. All images were

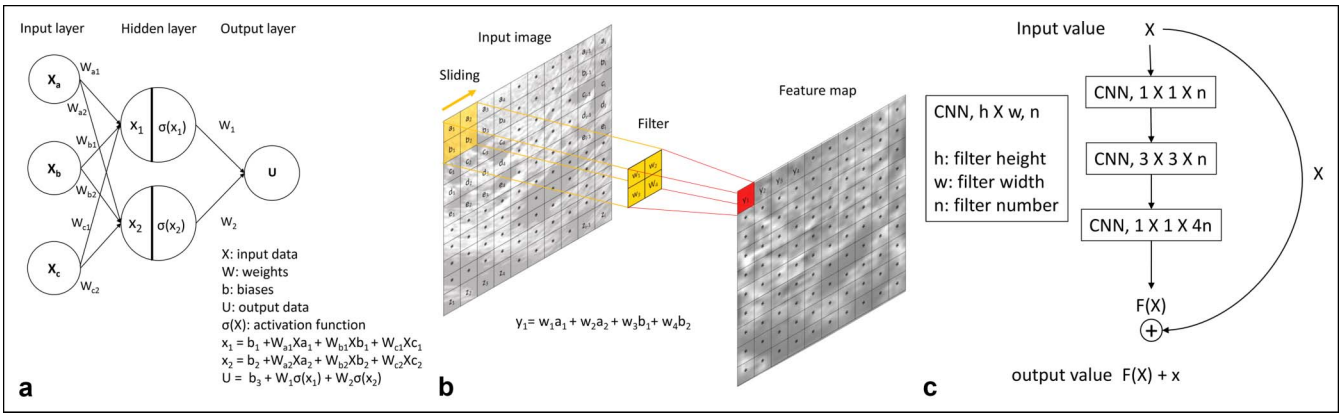

Figure 1. The basic concept of deep learning algorithm. (**a**) Layout of simple neural networks. A neural network consists of an input layer, a hidden layer, and an output layer. Input data (X_a, X_b, X_c) were converted to the values at the hidden layer (X_1, X_2) by applying a weight (W) to the input data, adding the bias (b), and passing to the activation function [$\sigma(X)$]. The values at the hidden layer (X_1, X_2) were then converted to the output data (U) using the same procedure ($U = b_3 + W_1\sigma(x_1) + W_2\sigma(x_2)$). (**b**) Layout of CNN. In the CNN, the filter size and number were freely selected, and each pixel value of the input images (a_1, a_2, b_1, b_2, etc.) was converted to the feature maps (y_1, y_2, y_3, etc.) by multiplying the filter weights (w_1, w_2, w_3, etc.) and sliding the filter over the input images (e.g., $y_1 = w_1a_1 + w_2a_2 + w_3b_1 + w_4b_2$). (**c**) Layout of residual block. In the residual block, there are 3 CNN layers, 2 filter sizes (1×1 and 3×3), 2 filter numbers (n or 4n), and shortcut connection between the CNN layers. The input value (x) was converted to the output value [$F(x)$] by CNN layers. The final output value of the residual block is $F(x) + x$. CNN, convolutional neural network.

randomly selected as test or training data, and all images of each patient were not divided into test and training data to prevent data leakage.

The primary end point was the accuracy of the diagnosis for the malignancy of IPMN *via* the AI value. The secondary end points were the accuracy of the diagnosis for the malignancy of IPMN *via* AI malignant probability, human preoperative diagnosis, conventional logistic regression analysis, and relative and absolute indications that were reported in the guidelines (8,9).

Statistical analyses

SPSS version 23.0 (SPSS, Chicago, IL) was used for all statistical analyses. All tests were 2 tailed, and $P < 0.05$ was considered

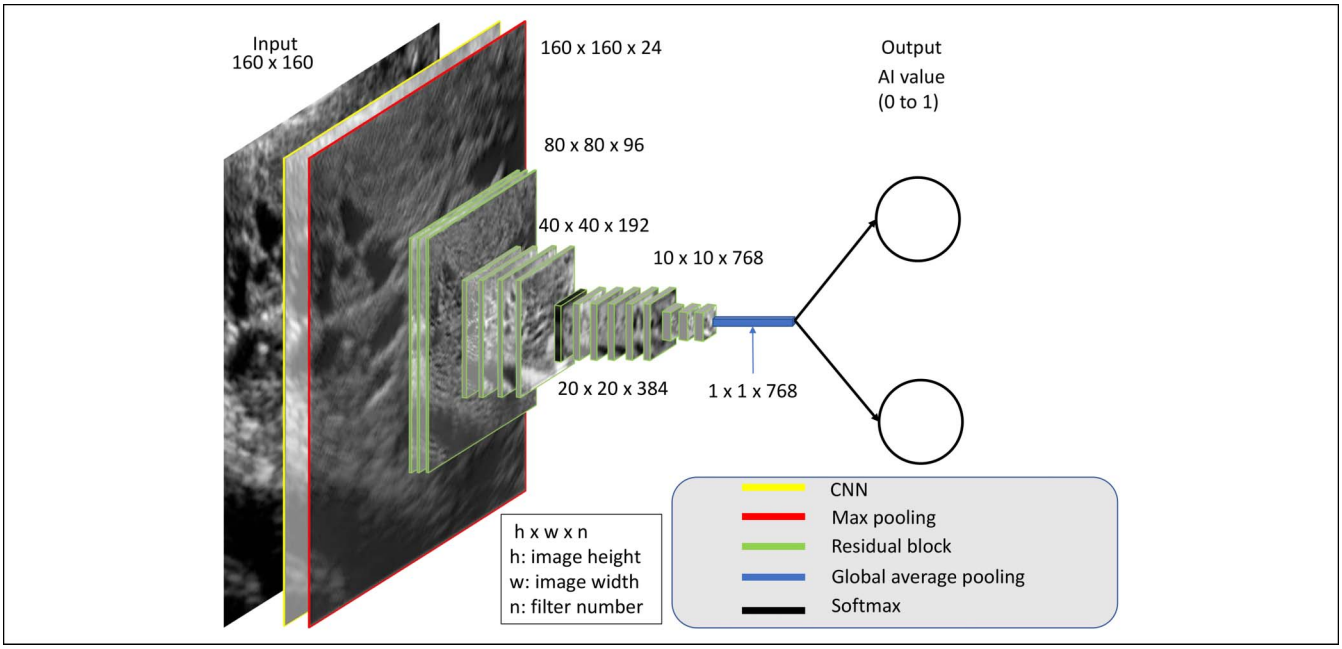

Figure 2. Layout of the deep learning algorithm. Data flow is from left to right: an image of EUS images (for example, benign intraductal papillary mucinous neoplasm) was used. All EUS images were trimmed to equally-sized squares (160×160 pixels). EUS images were then converted into levels on a gray scale (0–255) in each pixel. Subsequently, all EUS images were converted to mathematical information and were input to the algorithm. EUS images were processed by CNN, max pooling, and global average pooling layers. Swish activation functions were used for the hidden layers, whereas softmax functions were used for the output layers. Passing through the algorithm, image sizes gradually became reduced and filter numbers increased. As a result, the output value of deep learning was calculated as the predictive value of malignancy using AI (AI value: continuous variables from 0 to 1). AI, artificial intelligence; CNN, convolutional neural network; EUS, endoscopic ultrasonography.

statistically significant. Continuous variables were expressed as mean and SD or median and range. The Fisher exact test was used for categorical variables, and the Mann-Whitney *U* test was used for continuous variables. A receiver operating characteristic (ROC) curve was generated, and the area under the ROC curve (AUROC) was calculated to determine the cutoff value for the diagnosis of malignancy. AUROC accuracy was defined as low (0.5 to <0.7), moderate (0.7 to <0.9), or high (≥0.9). Cutoff values were determined to maximize the Youden index (sensitivity + specificity − 1), and sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy were calculated for these cutoff values. Using the cutoff values that were determined using ROC analysis, upper limit of normal in this institute, or the values that were defined in the guidelines, univariate relationships between the malignancy of IPMN and patients' characteristics including the image findings (cyst size, mural nodule size, and MPD diameter) and preoperative laboratory values (AMY, CEA, and CA19-9) were evaluated *via* logistic regression analysis. A multivariate logistic regression analysis was performed and included the variables that were significantly related to the malignancy of IPMN in the univariate analysis.

RESULTS

Among the 206 patients, 50 patients whose EUS images of IPMN were recorded in a digital format were fully investigated in this study. A total of 3,970 still images were collected from the 50 patients. Using the data augmentation technique, 508,160 still images were generated and fully investigated.

Characteristics of patients

The characteristics of patients are shown in Table 1. The final pathological diagnoses of IPMNs were benign ($n = 27$) and malignant ($n = 23$). The median age of all patients was 66 years (range, 18–81), and sex was distributed equally (men 50%). History of pancreatitis was observed in 5 patients (10%), whereas jaundice and new onset of diabetes mellitus were not noted in all patients. The surgical indications were suspected malignancy (84%), pancreatitis (10%), and malignancy of other organs (6%). The IPMNs were located in the head ($n = 32$), body ($n = 6$), and tail ($n = 12$). The IPMN types in all regions were branch duct ($n = 14$), main duct ($n = 6$), and mixed ($n = 30$). The median mural nodule size of malignant IPMNs was significantly higher than that of benign IPMNs ($P = 0.003$). However, histories of pancreatitis, AMY, CEA, CA19-9, MPD diameter, cyst size, and the ratio of growth rate (>5 mm/yr) were not significantly different between benign and malignant IPMNs.

Table 1. Patients' characteristics

	Benign IPMN, $n = 27$	Malignant IPMN, $n = 23$	*P* value
Age, yr, median (range)	69 (45–80)	71 (40–85)	0.626
Male sex, *n* (%)	13 (48)	12 (52)	1.000
History of pancreatitis, *n* (%)	3 (11)	2 (9)	1.000
Surgical indication			
Suspected malignancy	21 (78)	21 (91)	0.235
Pancreatitis	3 (11)	2 (9)	
Malignancy of other organs	3 (11)	0 (0)	
Location, *n* (%)			
Head	17 (63)	15 (65)	0.403
Body	2 (7)	4 (17)	
Tail	8 (30)	4 (17)	
Type, *n* (%)			
Branch duct	13 (48)	1 (4)	<0.001
Mixed	12 (44)	18 (78)	
Main duct	2 (7)	4 (17)	
Serum AMY, IU/L, median (range)	76 (36–297)	83 (54–290)	0.442
Serum CEA, ng/mL, median (range)	2.7 (0.7–35.4)	2.4 (0.9–6.1)	0.447
Serum CA19-9, U/mL, median (range)	14.4 (3.8–531.0)	13.3 (0.1–783.0)	0.189
Mural nodule size, mm, median (range)	0.0 (0.0–20.0)	9.0 (0.0–33.0)	0.003
MPD diameter, mm, median (range)	5.0 (2.0–20.0)	8.5 (2.0–18.5)	0.096
Cyst size, mm, median (range)	26.4 (1.0–52.0)	32.0 (1.0–50.0)	0.815
Growth rate (>5 mm/yr) n (%)	10 (37)	11 (48)	0.567

AMY, amylase; CA19-9, carbohydrate antigen 19-9; CEA, carcinoembryonic antigen; IPMN, intraductal papillary mucinous neoplasm; MPD, main pancreatic duct.

Diagnostic performance for the diagnosis of IPMN malignancy

The mean AI values (predictive value of IPMN malignancy in each image) of benign and malignant IPMNs were 0.104 ± 0.279 and 0.808 ± 0.367, respectively, and the AI value of malignant IPMNs was significantly higher than that of benign IPMNs ($P < 0.001$). The AUROC for ability to diagnose the malignancy of IPMN *via* the AI value was 0.91 (Figure 3). When an AI value of 0.49 was used as a cutoff point according to ROC analysis, the sensitivity, specificity, PPV, NPV, and accuracy were 81.5%, 90.1%, 86.5%, 86.2%, and 86.2%, respectively. The mean AI malignant probabilities (mean AI value of all images for each patient) of benign and malignant IPMNs were 0.109 ± 0.151 and 0.787 ± 0.227, respectively, and AI malignant probability of malignant IPMNs was significantly higher than that of benign IPMNs ($P < 0.001$). The AUROC for ability to diagnose the malignancy of IPMN *via* AI malignant probability was 0.98 and was significantly greater than the ability to diagnose the malignancy of IPMN *via* the mural nodule size (0.74, $P < 0.001$) and conventional logistic regression analysis (0.73, $P < 0.001$) (Figure 4). When AI malignant probability of 0.41 was used as a cutoff point (according to the ROC analysis), the sensitivity, specificity, PPV, NPV, and accuracy was 95.7%, 92.6%, 91.7%, 96.2%, and 94.0%, respectively. The diagnostic ability of AI malignant probability for predicting IPMN malignancy (accuracy: 0.94) was greater than human preoperative diagnosis (accuracy: 0.56), conventional logistic regression analysis (accuracy: 0.72), and relative and absolute indications that were reported in the guidelines (accuracy: 0.40–0.68) (Table 2).

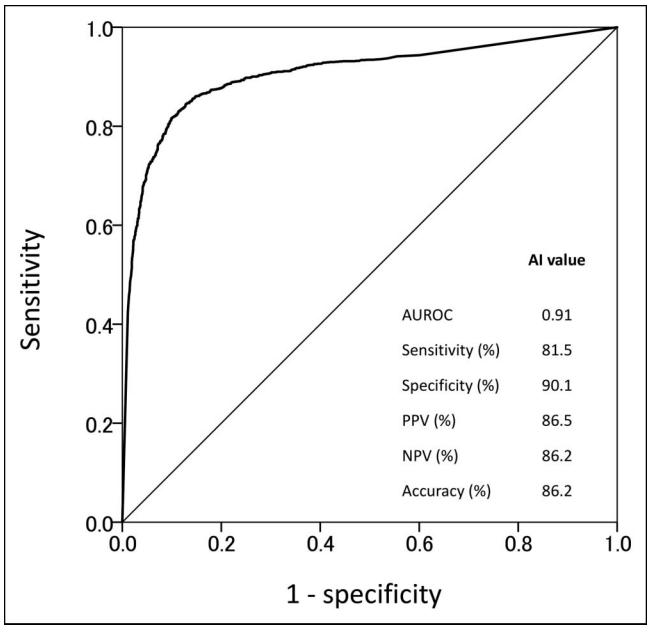

Figure 3. AUROC for ability to diagnose the malignancy of IPMN of the AI value. The output value of deep learning was calculated as the predictive value of malignant probability using AI (AI value: continuous variables from 0 to 1). The AUROC for ability to diagnose the malignancy of IPMN *via* the AI value was 0.91. AI, artificial intelligence; AUROC, area under the receiver operating characteristic curve; IPMN, intraductal papillary mucinous neoplasms.

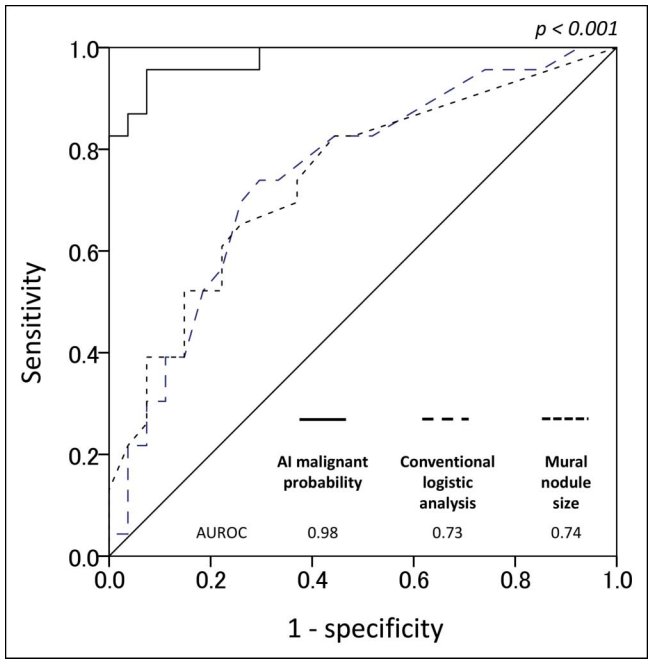

Figure 4. AUROC for ability to diagnose the malignancy of IPMNs of AI malignant probability, conventional logistic regression analysis, and the mural nodule size. The AUROC was 0.73 in conventional regression analysis, 0.74 in the mural nodule size, and 0.98 in AI malignant probability (the mean AI value of all images in each patient). AI, artificial intelligence; AUROC, area under the receiver operating characteristic curve; IPMN, intraductal papillary mucinous neoplasms.

Univariate and multivariate analyses of diagnostic performance for the IPMN malignancy

In univariate analysis, IPMN type (mixed or MPD type), serum CA19-9 ≥38 U/mL, mural nodule size ≥5 mm, and AI malignant probability ≥0.41 were significantly associated with the malignancy of IPMN (Table 3). In multivariate logistic regression analysis, AI malignant probability ≥0.41 was the only identified independent factor for the malignancy of IPMN with an odds ratio of 295.16 (95% confidence interval: 14.13–6,165.75, $P < 0.001$) (Table 3).

DISCUSSION

AI is a new technique for the objective evaluation of image information. In this study, we found that the AI value evaluated using the deep learning algorithm was significantly correlated with the malignancy of IPMN. Moreover, we found that the diagnostic performance of AI was higher than human diagnosis, conventional logistic regression analysis, and relative and absolute indications that were reported in the guidelines. In the international guidelines, several findings show well-known risks for malignancy, which have been used to assess the preoperative malignancy in numerous studies (3–7,9). However, the findings demonstrated inadequate performance, and several of them (caliber change of the pancreatic duct and wall thickness) had unclear objective criteria. Therefore, the determination of these findings tends to be subjective. In contrast to the diagnosis of IPMN, according to the guidelines, AI can objectively measure the malignancy before surgery by only using the EUS image. This study demonstrated that the assessment of malignancy of IPMN *via* AI was superior to all risk factors according to the guidelines. In the ROC curve analysis, the AUROC of AI malignant probability was greater than that of the mural nodule size, which was only significantly different between benign and malignant IPMNs in all risk factors of malignancy. Moreover, the accuracy of the AI malignant probability was greater than that of human preoperative diagnosis and conventional logistic regression analysis, indicating the superior predictive ability of AI. In multivariate analysis, which included various putative risk factors of malignancy, only AI malignant probability was identified as an independent risk factor for IPMN malignancy. These results indicated that AI is a useful tool for objective diagnosis of malignancy of IPMN.

This study is the first to attempt to diagnose the malignancy of IPMN *via* AI. AI has been used for the diagnosis of several diseases (eye and skin cancer, breast tumors, and colorectal polyps) using image information such as computed tomography, magnetic resonance imaging, and endoscopic images (15–18,30,31). The diagnostic performance of AI was reported to be higher than that of human diagnosis. In pancreatic diseases, AI was used for the differential diagnosis of pancreatic tumors using endoscopic elastography images and contrast-enhanced ultrasonography images (32,33). In these reports, the diagnostic performance of AI was higher than that of only EUS findings. However, the algorithm that was used in these studies was multilayer perceptron (MLP). The input for MLP is a numeric value, such as laboratory values and image findings, that was speculated to be important for diagnostic treatment. By contrast, the input information of the deep learning algorithm that was used in this study is image information itself. Moreover, the diagnostic performance of the deep learning algorithm is superior to that of MLP. Future prospective and consecutive studies are warranted to evaluate the diagnostic performance among only image findings, MLP, and CNN.

Table 2. Diagnostic ability of AI malignant probability to diagnose the malignancy of IPMN

	Cutoff point	Sensitivity	Specificity	PPV	NPV	Accuracy
AI malignant probability	0.41	0.957	0.926	0.917	0.962	0.940
Human preoperative diagnosis	+	0.957	0.222	0.512	0.857	0.560
Logistic regression analysis	65	0.739	0.704	0.680	0.760	0.720
Mural nodules	5 mm	0.739	0.630	0.630	0.739	0.680
MPD diameter	10 mm	0.435	0.704	0.556	0.594	0.580
Cyst size	40 mm	0.217	0.778	0.455	0.538	0.520
CA19-9	37U/mL	0.087	0.667	0.182	0.462	0.400
Growth rate	>5 mm/yr	0.478	0.630	0.524	0.586	0.560
Pancreatitis	+	0.261	0.778	0.500	0.553	0.540

AI, artificial intelligence; CA19-9, carbohydrate antigen 19-9; IPMN, intraductal papillary mucinous neoplasm; MPD, main pancreatic duct; NPV, negative predictive value; PPV, positive predictive value.

The current study had several limitations. First, it was a retrospective single-center study. Prospective investigations conducted at multiple institutions are necessary for validating the results obtained in the current study. Second, only a small sample size was included. Therefore, internal validation (10-fold cross-validation) was used to evaluate the diagnostic performance for the malignancy of IPMN because we could not collect enough patients to perform AI when patients were separated into groups for training and test data. When a diagnostic performance is evaluated, test data should be separated from the training data, and the number of both training and test data should be large. AI techniques, such as data augmentation and transfer learning, have been recently developed to overcome this limitation (34). Using these new techniques, AI can achieve an adequate diagnostic performance in small sample sizes. In this study, the data augmentation technique was used. Therefore, over 500,000 images, which were enough for AI, were generated. However, to evaluate the real diagnostic performance of AI, more patients from multiple centers are needed in future studies. Third, only surgical cases were included in this study. In clinical practice, most patients with IPMN undergo surveillance rather than surgical resection. Therefore, bias may have occurred. However, several regions that were not suspected as malignant were included in this study because the surgical indication was not only suspected malignancy but also pancreatitis and malignancy of other organs, which may have reduced the bias.

Table 3. Univariate/multivariate logistic regression analysis of relationship between malignancy of IPMN and patients' characteristics

	Univariable		Multivariable	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value
Age (≥70 yr)	1.89 (0.61–5.83)	0.268		
Sex (male)	0.85 (0.28–2.59)	0.777		
Pancreatitis	1.23 (0.34–4.53)	0.750		
Location (body or tail)	1.09 (0.35–3.43)	0.879		
IPMN type (mixed or MPD)	20.49 (2.40–173.89)	0.006	24.78 (0.57–1,084.64)	0.096
Serum AMY (≥126 IU/L)	1.51 (0.32–7.17)	0.600		
Serum CEA (≥5.1 ng/mL)	1.83 (0.30–11.02)	0.511		
Serum CA19-9 (≥38 U/mL)	5.25 (1.00–27.5)	0.050	1.33 (0.04–41.97)	0.873
Mural nodule size (≥5 mm)	4.82 (1.43–16.23)	0.011	4.53 (0.29–70.29)	0.281
MPD diameter				
≥5 mm, <10 mm	0.26 (0.05–1.33)	0.106		
≥10 mm	2.48 (0.61–10.06)	0.205		
Cyst size (≥40 mm)	0.97 (0.25–3.73)	0.967		
Growth rate (>5 mm/yr)	1.56 (0.50–4.83)	0.442		
Human preoperative diagnosis	6.29 (0.69–56.72)	0.101		
AI malignant probability (≥0.41)	275.00 (23.31–3,244.31)	<0.001	295.16 (14.13–6,165.75)	<0.001

AI, artificial intelligence; AMY, amylase; CA19-9, carbohydrate antigen 19-9; CEA, carcinoembryonic antigen; CI, confidence interval; IPMN, intraductal papillary mucinous neoplasm; MPD, main pancreatic duct; OR, odds ratio.

In conclusion, the AI value measured *via* AI in patients who had malignant IPMNs was higher than that in patients who had benign IPMNs, and the accuracy *via* the AI value was 86.2%. Among various clinical characteristics, the AI malignant probability was the only independent diagnostic factor that significantly predicted the malignancy of IPMNs. The use of AI is recommended for objectively assessing the preoperative malignancy of IPMNs.

CONFLICTS OF INTEREST

Guarantor of the article: Takamichi Kuwahara, MD, PhD.
Specific author contributions: All authors had access to the study data and had reviewed and approved the final manuscript. T.K.: study concept and design, acquisition of data, analysis and interpretation of data, drafting of the manuscript, and statistical analysis. K.H., N.M., N.O., S.M., M.O., Y.K., H.K., K.T., S.O., M.I., T.T., and M.T.: acquisition of data and critical revision of the manuscript for important intellectual content. Y.N.: study supervision.
Financial support: This work was supported by JSPS KAKENHI Grant Number JP 18K15769; the work was independent of it.
Potential competing interests: None.

Study Highlights

WHAT IS KNOWN

- ✓ Difficult preoperative diagnosis of IPMN malignancy.
- ✓ Deep learning provides a high-performance prediction and has been applied for medical diagnosis.

WHAT IS NEW HERE

- ✓ The accuracy *via* AI for malignancy diagnosis of IPMNs was 86.2%.
- ✓ AI *via* deep learning increased the diagnostic accuracy in detecting malignancy of IPMNs.

TRANSLATIONAL IMPACT

- ✓ AI malignant probability was the only independent diagnostic factor that significantly predicted the IPMN malignancy.
- ✓ AI diagnosis for the malignancy of IPMNs was more accurate than human diagnosis and conventional diagnosis methods.

REFERENCES

1. Brosens LAA, Hackeng WM, Offerhaus GJ, et al. Pancreatic adenocarcinoma pathology: Changing "landscape". J Gastrointest Oncol 2015;6:358–74.
2. Moris D, Damaskos C, Spartalis E, et al. Updates and critical evaluation on novel biomarkers for the malignant progression of intraductal papillary mucinous neoplasms of the pancreas. Anticancer Res 2017;37:2185–94.
3. Tanaka M, Fernández-del Castillo C, Adsay V, et al. International consensus guidelines 2012 for the management of IPMN and MCN of the pancreas. Pancreatology 2012;12:183–97.
4. Tanaka M, Fernández-del Castillo C, Kamisawa T, et al. Revisions of international consensus Fukuoka guidelines for the management of IPMN of the pancreas. Pancreatology 2017;17:738–53.
5. Yu S, Takasu N, Watanabe T, et al. Validation of the 2012 Fukuoka consensus guideline for intraductal papillary mucinous neoplasm of the pancreas from a single institution experience. Pancreas 2017;46:936–42.
6. Seo N, Byun JH, Kim JH, et al. Validation of the 2012 International Consensus Guidelines using computed tomography and magnetic resonance imaging: Branch duct and main duct intraductal papillary mucinous neoplasms of the pancreas. Ann Surg 2016;263:557–64.
7. Heckler M, Michalski CW, Schaefle S, et al. The Sendai and Fukuoka consensus criteria for the management of branch duct IPMN: A meta-analysis on their accuracy. Pancreatology 2017;17:255–62.
8. European Study Group on Cystic Tumours of the Pancreas. European evidence-based guidelines on pancreatic cystic neoplasms. Gut 2018;67:789–804.
9. Shimizu Y, Hijioka S, Hirono S, et al. New model for predicting malignancy in patients with intraductal papillary mucinous neoplasm. Ann Surg 2018. [Epub ahead of print November 29, 2018.]
10. Shimizu Y, Yamaue H, Maguchi H, et al. Predictors of malignancy in intraductal papillary mucinous neoplasm of the pancreas: Analysis of 310 pancreatic resection patients at multiple high-volume centers. Pancreas 2013;42:883–8.
11. Gemenetzis G, Bagante F, Griffin JF, et al. Neutrophil-to-lymphocyte ratio is a predictive marker for invasive malignancy in intraductal papillary mucinous neoplasms of the pancreas. Ann Surg 2017;266:339–45.
12. Ngamruengphong S, Bartel MJ, Raimondo M. Cyst carcinoembryonic antigen in differentiating pancreatic cysts: A meta-analysis. Dig Liver Dis 2013;45:920–6.
13. Takano S, Fukasawa M, Kadokura M, et al. Next-generation sequencing revealed TP53 mutations to be malignant marker for intraductal papillary mucinous neoplasms that could be detected using pancreatic juice. Pancreas 2017;46:1281–7.
14. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–44.
15. 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;318:2211–23.
16. 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;316:2402–10.
17. 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;318:2199–210.
18. Esteva A, Kuprel B, Novoa RA, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542:115–8.
19. World Medical Association Inc. Declaration of Helsinki: Ethical principles for medical research involving human subjects. J Indian Med Assoc 2009;107:403–5.
20. Krizhevsky A., Sutskever I., Hinton G. ImageNet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. 2012;25:1097–1105.
21. Szegedy C., Liu W, Jia Y, et al. Going deeper with convolutions. arXiv preprint arXiv. 2014;1409:4842.
22. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv. 2014:1409.1556.
23. He K., Zhang X., Ren, et al. Deep residual learning for image recognition. arXiv preprint arXiv. 2015;1512:03385.
24. Ramachandran P, Zoph B, Le Q. Searching for activation functions. arXiv Preprint arXiv.2017:1710.05941v2.
25. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv. 2015:1502.03167.
26. Huang G, Sun Y, Liu Z, et al. Deep networks with stochastic depth. https://arxiv.org/abs/1603.09382. Accessed on April 29, 2019.
27. Yao Y, Rosasco L, Caponnetto A. On early stopping in gradient descent learning. Constr Approx 2007;26:289–315.
28. Fadaee M, Bisazza A, Monz C. Data augmentation for low-resource neural machine translation. arXiv preprint arXiv. 2017:1705.00440.
29. Zhong Z, Zheng L, Kang G, et al. Random erasing data augmentation. arXiv preprint arXiv 2017:1708.04896v2.
30. 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.

31. Chen PJ, Lin MC, Lai MJ, et al. Accurate classification of diminutive colorectal polyps using computer-aided analysis. Gastroenterology 2018; 154:568–75.
32. Săftoiu A, Vilmann P, Gorunescu F, et al. Efficacy of an artificial neural network-based approach to endoscopic ultrasound elastography in diagnosis of focal pancreatic masses. Clin Gastroenterol Hepatol 2012;10:84–90.e1.
33. Săftoiu A, Vilmann P, Dietrich CF, et al. Quantitative contrast-enhanced harmonic EUS in differential diagnosis of focal pancreatic masses (with videos). Gastrointest Endosc 2015;82:59–69.
34. Pan SJ, Yang Q. A survey on transfer learning. IEEE T Knowl Data En 2010;22:1345–59.