

Original Article

Diagnosing gangrenous cholecystitis on computed tomography using deep learning: A preliminary study

Yoichi Okuda,^{1,2} Tsukasa Saida,³ Keigo Morinaga,⁴ Arisa Ohara,⁴ Akihiro Hara,¹ Shinji Hashimoto,¹ Shinji Takahashi,¹ Tomoaki Goya,¹ and Nobuhiro Ohkohchi²

¹Department of Surgery, Koyama Memorial Hospital, Kashima, Japan, ²Department of Surgery, Mitochuo Hospital, Mito, Japan, ³Department of Radiology, Faculty of Medicine, University of Tsukuba, Tsukuba, Japan, and ⁴Department of Radiology, Koyama Memorial Hospital, Kashima, Japan

Aim: To compare deep learning and experienced physicians in diagnosing gangrenous cholecystitis using computed tomography images and explore the feasibility of diagnostic assistance for acute cholecystitis requiring emergency surgery.

Methods: This retrospective study included 25 patients with pathologically confirmed gangrenous cholecystitis and 129 patients with noncomplicated acute cholecystitis who underwent computed tomography between 2016 and 2021 at two institutions. All available computed tomography images at the time of the initial diagnosis were used for the analysis. A deep learning model based on a convolutional neural network was trained using 1,517 images of 112 patients (18 patients with gangrenous cholecystitis and 94 patients with acute cholecystitis) and tested with 68 images of 42 patients (seven patients with gangrenous cholecystitis and 35 patients with acute cholecystitis). Three blinded, experienced physicians independently interpreted the test images. The sensitivity, specificity, accuracy, and area under the receiver operating characteristic curve were compared between the convolutional neural network and the reviewers.

Results: The convolutional neural network (sensitivity, 0.70; 95% confidence interval [CI], 0.44–0.87, specificity, 0.93; 95% CI, 0.88–0.96, accuracy, 0.89; 95% CI, 0.81–0.95, area under the receiver operating characteristic curve, 0.84; 95% CI, 0.68–1.00) had achieved a better diagnostic performance than the reviewers (ex. sensitivity, 0.55; 95% CI, 0.30–0.77, specificity, 0.67; 95% CI, 0.62–0.71, accuracy, 0.65; 95% CI, 0.57–0.72, area under the receiver operating characteristic curve, 0.63; 95% CI, 0.44–0.82; $P = 0.048$ for area under the receiver operating characteristic curve versus convolutional neural network).

Conclusions: Deep learning had a better diagnostic performance than experienced reviewers in diagnosing gangrenous cholecystitis and has potential applicability for assisting in identifying indications for emergency surgery in the future.

Key words: Acute cholecystitis, artificial intelligence, computed tomography, convolutional neural network

INTRODUCTION

ACUTE CHOLECYSTITIS (AC) is a common disease that causes acute abdominal pain. The Tokyo Guidelines 2018 (TG18) are currently used to assess the severity of AC.¹ Gangrenous cholecystitis (GC) is listed as a condition for the diagnosis of moderate cholecystitis in the TG18 and is defined as the histologic presence of acute transmural

inflammation and ischemic necrosis of the gallbladder wall.¹ Because delayed diagnosis can lead to sepsis, gallbladder perforation, and intra-abdominal abscess and fistula formation, it is important to distinguish between GC, which often requires urgent cholecystectomy, and noncomplicated AC, which can be treated conservatively with antibiotics and percutaneous image-guided drainage using preoperative imaging.^{2,3}

Ultrasonography (US) is the most common and optimal examination when AC is clinically suspected. However, computed tomography (CT) is widely used for the evaluation of gallbladder disease and allows rapid imaging for patients presenting with nondemonstrable areas during the initial workup, and the TG18 state that contrast-enhanced CT is superior to US for diagnosing GC.^{1,4} Previous studies have evaluated individual CT signs for the diagnosis of

Corresponding: Tsukasa Saida, MD, Department of Radiology, Faculty of Medicine, University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8575, Ibaraki, Japan. E-mail: saida_sasaki_tsukasa@md.tsukuba.ac.jp.

Received 14 Jun, 2022; accepted 17 Aug, 2022

Funding information

No funding information provided.

GC²⁻⁵; it has been reported that distention, mural striation, and decreased gallbladder wall enhancement can significantly distinguish GC from noncomplicated AC.² However, surgeons cannot always easily determine the indications for urgent surgery using CT in an emergency.

Deep learning is a main area of focus in medicine today and has already been applied to multiple domains in this field. Convolutional neural networks (CNNs) are considered promising tools for diagnostic imaging; several CNNs have recently been constructed and achieved excellent performance in image classification.^{6,7}

As a preliminary study for diagnosing surgically indicated cholecystitis, we presented a CNN for distinguishing GC from noncomplicated AC using CT images and compared its diagnostic performance with the interpretations of experienced physicians.

METHODS

Patients

THE PROTOCOL FOR this research project was approved by the Ethics Committees of Koyama Memorial Hospital and Mitochuo Hospital (approval numbers 21015 and 3-002) and written informed consent was waived. The inclusion criteria were as follows: (i) clinically definite AC according to the TG18; (ii) CT images obtained at the time of diagnosis between January 2016 and October 2021 at the two institutions; (iii) pathologically proven GC (defined as the GC group); and (iv) pathologically proven nonGC or mild cholecystitis not requiring surgery (defined as the AC group). The exclusion criteria were as follows: (i) mild cholecystitis not requiring surgery and not followed up for more than a year; and (ii) moderate or severe cholecystitis that could not be treated surgically because of other complications. The severity of AC was diagnosed according to the TG18. GC was pathologically diagnosed by full-thickness necrosis or ulceration of the gallbladder wall.^{8,9} A flowchart of the patient selection process is shown in Figure 1.

CT acquisition

Because of the retrospective nature of this study and the inclusion of patients from two different institutions, the CT scan protocol was not standardized. CTs were performed with multidetector CT scanners (LightSpeed 16 or LightSpeed VCT 64; GE Healthcare, Milwaukee, WI). The CT parameters were as follows: tube voltage, 120–140 kVp; gantry rotation time, 0.5 s; and noise index, 9.0 at a 5-mm slice thickness. Contrast-enhanced images were acquired with a bolus injection of intravenous (IV) iohexol

(Omnipaque 350, Nycomed Amersham, Princeton, NJ) at a dose of 1.7 mL/kg body weight. Images with a 5.0-mm slice thickness were obtained.

Of the 154 patients, 150 underwent noncontrast CT, 4 underwent contrast-enhanced CT only, and 42 underwent both noncontrast and contrast-enhanced CT. In addition, 18 patients underwent dynamic contrast CT.

Dataset

To create a dataset, only the slices in which the lumen of the gallbladder was visualized were extracted from all CT images, as per the consensus of two surgeons (Y.O., A.H.).

A total of 154 patients (25 patients with GC and 129 patients with AC) were randomly assigned to the training and test groups in equal proportions. The training set consisted of 1,517 images from 112 patients (354 images from 18 patients in the GC group and 1,163 images from 94 patients in the AC group). The test set consisted of the central slices of the gallbladder, resulting in 68 images from 42 patients (11 images of 7 patients in the GC group and 57 images of 35 patients in the AC group).

In this study, the software used was unable to handle Digital Imaging and Communications in Medicine (DICOM) format images, so they were converted to joint photographic experts group (JPEG) format images after adjusting the window level and width to ensure appropriate interpretation using the Centricity Universal Viewer (GE Healthcare, Chicago, IL). Next, the margins were automatically cropped, and the images were automatically resized to 240 × 240 pixels using XnConvert (Gougelet Pierre-Emmanuel, Reims, France).

Deep learning using CNNs

Deep learning was performed on a deep station entry (UEI, Tokyo, Japan) with a GeForce RTX 2080Ti graphics processing unit (NVIDIA, Delaware, CA), a Core i7-8700 central processing unit (Intel, Santa Clara, CA), and deep learning software Deep Analyzer (GHELIA, Tokyo, Japan).

The conditions for deep learning were optimized from ablation and comparative studies of previous research. A CNN with the Xception architecture was used for deep learning. The Xception architecture is characterized as having depthwise separable convolutions that enable the use of model parameters more efficiently than previous CNN architectures.¹⁰ ImageNet, which comprises natural images as pre-trained data, was used for pretraining.¹¹ Adam was selected as the optimizer algorithm. The learning rate was set to 0.0001, and horizontal flipping, rotation (4.5°), shearing (0.05), and zooming (0.05) were automatically used to augment the data. The training and validation ratio was set

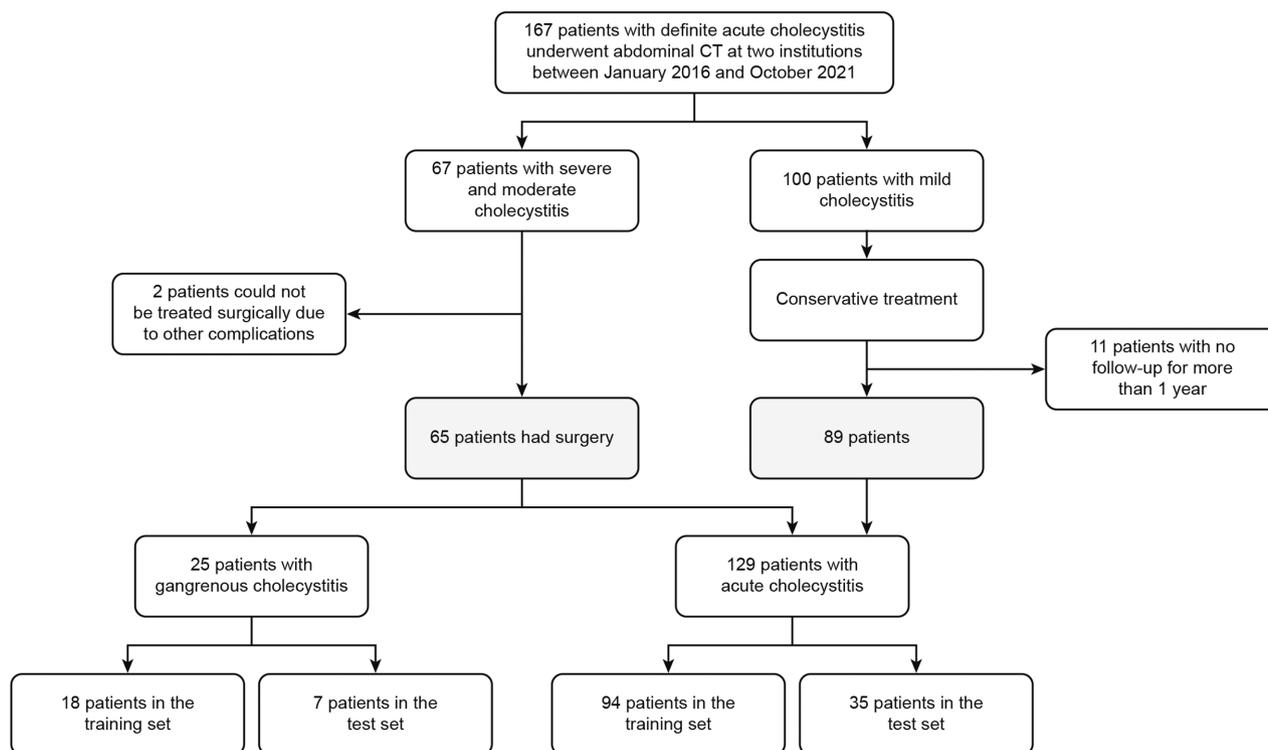


Fig. 1. Flowchart for the patient selection process.

to 4:1, and 50 epochs were used for the training. The batch size was automatically selected by Deep Analyzer to fit into the GPU memory.

Radiologist interpretation

An experienced surgeon (S.T.) and two experienced radiologists (K.M. and A.O.), with 26, 28, and 19 years of experience in interpreting CT independently reviewed the test images in random order. They evaluated each image by assigning confidence levels to the presence of GC using a six-point scale (0, definitely AC; 0.2, probably AC; 0.4, possibly AC; 0.6, possibly GC; 0.8, probably GC; 1, definitely GC). The reviewers were blinded to the pathological and clinical findings of the patients.

Statistical analysis

The age, sex, white blood cell (WBC) count, C-reactive protein (CRP) level, and time interval between CT and surgery for each group were compared using the Mann–Whitney U test and the χ^2 test.

For the reviewers' interpretations, 0.0–0.4 was treated as AC, whereas 0.6–1.0 was treated as GC. For the CNN, the

probability for GC was output as a continuous number from 0 (AC) to 1 (GC): 0.00–0.49 was considered AC, whereas 0.50–1.00 was considered GC. These AC or GC determinations were used to calculate the sensitivity, specificity, and accuracy of the GC diagnosis. Receiver operating characteristic (ROC) curve analysis was performed to assess the diagnostic performance of the CNN and the reviewers. Moreover, the areas under the receiver operating characteristic curve (AUC) and their 95% confidence intervals (CI) were compared between the CNN and the reviewers,¹² and significant differences were estimated.¹³

Interobserver agreement for the presence or absence of GC was also assessed using κ statistics.¹⁴

All statistical analyses were performed using SPSS software (SPSS Statistics 27.0; IBM, New York, NY). Statistical significance was set at $P < 0.05$.

RESULTS

A TOTAL OF 154 patients (mean age, 75 years; age range, 25–108 years) were evaluated across the datasets. Table 1 shows the patients' characteristics and the number of image slices demonstrating GC and AC lesions. Although, GC was significantly more common in men

Table 1. Characteristics of the patients and imaging data

	Training data			Testing data		
	Gangrenous	Acute	All	Gangrenous	Acute	All
Patients (n)	18 (M:F = 12:6)	94 (M:F = 53:41)	112 (M:F = 65:47)	7 (M:F = 6:1)	35 (M:F = 22:13)	42 (M:F = 28:14)
Age						
Mean ± standard deviation (year)	76 ± 12	75 ± 15	75 ± 15	71 ± 14	72 ± 13	73 ± 13
Range (year)	50–95	30–108	30–108	44–85	25–87	25–87
Images						
All (n/slices)	18/354	94/1163	112/1517	7/11	35/57	42/68
Noncontrast CT (n/slices)	18/223	92/770	110/993	6/6	34/34	40/40
Portal phase contrast CT (n/slices)	6/101	22/219	28/320	3/3	15/15	18/18
Arterial phase contrast CT (n/slices)	2/30	10/116	12/146	1/1	5/5	6/6
Equilibrium phase contrast CT (n/slices)	0	7/58	7/58	1/1	3/3	4/4

Abbreviations: F, female; M, male.

($P = 0.049$), there were no significant differences in patient age between the GC and AC groups ($P = 0.642$). The median values of WBC count and CRP were $14.674/\text{mm}^3$ and 15.13 mg/dL for GC and $10.890/\text{mm}^3$ and 9.37 mg/dL for AC, respectively; both values were significantly higher in patients with GC ($P = 0.005$ and 0.010 , respectively). The median time interval between CT and surgery for patients with GC was 2 days (range, 0–19 days), whereas for patients with AC was 19 days (range, 0–77 days), and the difference was statistically significant ($P = 0.036$).

Table 2 lists the diagnostic performance of the CNN versus that of the reviewers, and the corresponding ROC curves are shown in Figure 2. The results of the interpretation are shown in Table S1. The CNN showed the highest diagnostic performance with an AUC of 0.84, and its specificity, accuracy, and AUC were better than those of the three reviewers; the AUC of the CNN was significantly better than the surgeon and radiologist 2. Histograms of the CNN and reviewers, and calibration plots of the CNN are shown in Figures S1 and S2. The results of the CNN trained on non-contrast CT images only and contrast-enhanced CT images only are presented as Figure S3 for reference.

Figures 3 and 4 show the test images of two cases. Figure 3 shows a contrast-enhanced CT image of GC. Obscuration of the gallbladder wall can be recognized, and the CNN and all reviewers were able to correctly diagnose GC. Figure 4 shows a noncontrast CT image of noncomplicated AC, for which only the CNN made a correct diagnosis.

Table 3 shows the interobserver agreement between the assessments of the CNN and the three reviewers. The κ values between the CNN and the reviewers ranged from 0.04 to 0.17, indicating slight agreement and low consistency. Among the reviewers, the κ values ranged from 0.33 to 0.74, indicating fair to substantial agreement that varied widely.

DISCUSSION

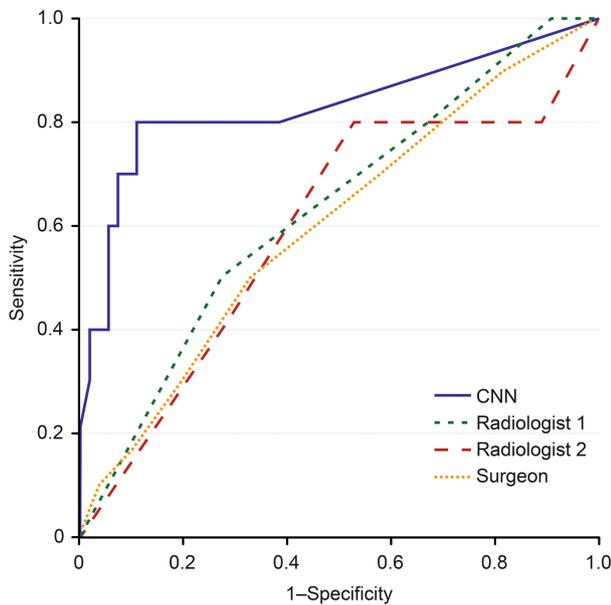
THIS STUDY PRESENTED a CNN for diagnosing GC requiring urgent surgery using CT images, which demonstrated a diagnostic performance that was better than that of experienced physicians.

In recent years, numerous applications of deep learning have been proposed in the medical field, several which have been reported in the gallbladder. Kim *et al.*¹⁵, Jeong *et al.*¹⁶, and Jang *et al.*¹⁷ evaluated the diagnostic performance of deep learning in differentiating polypoid lesions using US images. Loukas *et al.*¹⁸ proposed a deep learning approach for the assessment of gallbladder wall vascularity from images of laparoscopic cholecystectomy. Pang *et al.*¹⁹ developed a deep learning method for gallstone recognition using CT images. However, to the best of our knowledge, this is

Table 2. Sensitivity, specificity, and AUC of the convolutional neural network

Interpreter	Sensitivity	95% CI	Specificity	95% CI	Accuracy	95% CI	AUC	95% CI	P value for AUC (vs CNN)
CNN	0.70	0.44–0.87	0.93	0.88–0.96	0.89	0.81–0.95	0.84	0.68–1.00	—
Surgeon	0.55	0.30–0.77	0.67	0.62–0.71	0.65	0.57–0.72	0.63	0.44–0.82	0.048*
Radiologist 1	0.45	0.23–0.70	0.72	0.68–0.77	0.68	0.60–0.76	0.62	0.45–0.79	0.085
Radiologist 2	0.82	0.55–0.95	0.47	0.42–0.50	0.53	0.44–0.57	0.62	0.43–0.82	0.018*

Abbreviations: AUC, area under the receiver operating characteristic curve; CI, confidence interval; CNN, convolutional neural network.
* $P < 0.05$.

**Fig. 2.** Receiver operating characteristic curves for the performance of the convolutional neural network and the reviewers.

the first study to evaluate AC in CT images using deep learning. Notably, the images used in this study were not cropped images of the gallbladder alone.

GC exhibits specific findings on contrast-enhanced CT, including irregular thickening and poor contrast enhancement of the gallbladder wall, increased density of fatty tissue around the gallbladder, gas in the gallbladder lumen or wall, membranous structures within the lumen, and perigallbladder abscesses.¹ Among these CT signs, Chang *et al.*² reported that a short-axis diameter above 4.0 cm, mural striation (alternating areas of low and high attenuation of the wall), and decreased gallbladder wall enhancement were significant signs for distinguishing GC from AC. Of course, it is generally considered impossible to differentiate between GC and AC with only noncontrast CT images. However, CT for acute abdomen is usually performed urgently, and the

patient's situation does not always allow consent for contrast CT. This study population was relatively old and many of them had renal dysfunction, and the percentages of noncontrast CT images in the training (65%) and testing sets (59%) were quite high. This may be one of the reasons why the reviewers' diagnostic performance for GC was poor. There are reports that the hyperdense gallbladder wall-lumen sign on noncontrast CT images is useful in diagnosing GC^{20–22}; therefore, it may not be impossible to diagnose GC with noncontrast CT images. However, a retrospective review of our test cases showed that three of six patients (50%) in the GC group and four of 34 patients (12%) in the AC group had signs of a hyperdense gallbladder wall-lumen, limiting the diagnostic performance of physicians only on noncontrast CT images. Although CNN's decision-making process is black-box, the low agreement between the CNN and the reviewers suggests that CNN may have different criteria. Either way, a reading support system should desirably be able to cope with adverse conditions such as noncontrast CT images, and the good diagnostic performance of our CNN on test data with a high percentage of noncontrast CT images is promising for its future prospects.

Our preliminary study had several limitations. First, the study population was small, especially in the GC group. Second, because the test images were intentionally selected, selection bias was unavoidable. Third, the AC group included patients with lesions that were not pathologically confirmed. Fourth, the image data were not uniform. However, the presence of variance in the image data is desirable for creating a more generalizable CNN. Beyond this preliminary study, we aim to develop a clinically applicable CNN by using DICOM data, incorporating clinical information such as WBC counts and CRP—that showed significant differences in this study—and using images taken by various CT systems.

In conclusion, deep learning showed good diagnostic performance with CT images in the diagnosis of GC, despite

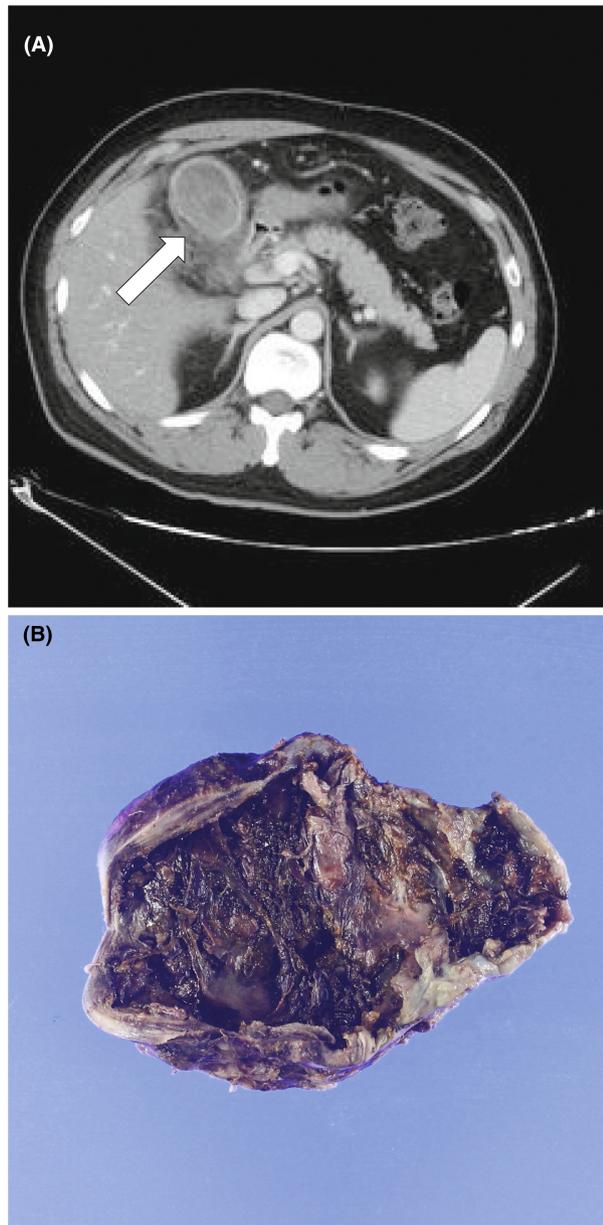


Fig. 3. Gangrenous cholecystitis in a 44-year-old man. (A) Test image: the CNN and all reviewers successfully diagnosed the gangrenous cholecystitis (confidence value for gangrenous cholecystitis: CNN = 86%, surgeon = 80%, radiologist 1 = 60%, radiologist 2 = 100%). On the right side, the gallbladder wall (arrow) is indistinct, and the surrounding fatty tissue is highly opacified. (B) Resected specimen: the gallbladder wall is reddish-brown in color, and the mucosal surface is strongly devitalized, indicating gangrenous cholecystitis. CNN, convolutional neural network.

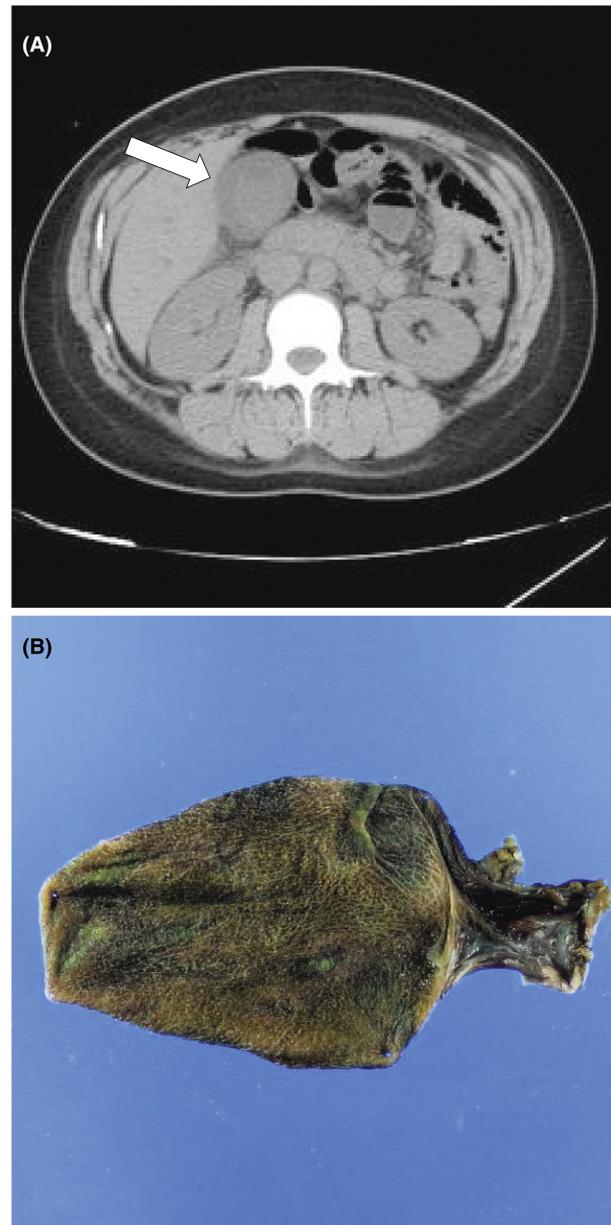


Fig. 4. Noncomplicated acute cholecystitis in a 25-year-old woman. (A) Test image: only the CNN successfully diagnosed the noncomplicated acute cholecystitis (confidence value for gangrenous cholecystitis: CNN = 0%, surgeon = 60%, radiologist 1 = 60%, radiologist 2 = 80%). The high-density gallbladder wall-lumen sign (arrow) appears to be present, although the wall itself is difficult to assess on the noncontrast CT image. High opacity is seen around the gallbladder neck. (B) Resected specimen: the mucosal surface of the gallbladder wall is smooth and well preserved, with no color abnormalities, indicating noncomplicated acute cholecystitis. CNN, convolutional neural network.

Table 3. ICC between the C and the reviewers

	Reviewers	ICC
CNN vs reviewers	Surgeon	0.04
	Radiologist 1	0.05
	Radiologist 2	0.17
Between reviewers	Surgeon vs radiologist 1	0.74
	Surgeon vs radiologist 2	0.43
	Radiologist 1 vs radiologist 2	0.33

Abbreviations: CNN, convolutional neural network; ICC, interobserver agreement.

noncontrast CT images comprising more than half of the images in this study. In the future, deep learning may also be used to make decisions for emergency surgery for AC.

ACKNOWLEDGEMENTS

THE AUTHORS THANK R. Urakawa and A. Odagiri for technical assistance with the experiments.

DISCLOSURE

APPROVAL OF THE research protocol with approval number and committee name: Not applicable.

Informed consent: Written informed consent was waived.

Registry and the registration no. of the study/trial: no. 21015 and no. 3-002.

Animal studies: Not applicable.

Conflicts of interest: None declared.

REFERENCES

- Yokoe M, Hata J, Takada T *et al.* Tokyo Guidelines 2018: diagnostic criteria and severity grading of acute cholecystitis (with videos). *J. Hepatobiliary Pancreat. Sci.* 2018; 25: 41–54.
- Chang WC, Sun Y, Wu EH *et al.* CT findings for detecting the presence of gangrenous ischemia in cholecystitis. *AJR Am. J. Roentgenol.* 2016; 207: 302–9.
- Fuks D, Mouly C, Robert B, Hajji H, Yzet T, Regimbeau JM. Acute cholecystitis: preoperative CT can help the surgeon consider conversion from laparoscopic to open cholecystectomy. *Radiology* 2012; 263: 128–38.
- van Randen A, Laméris W, van Es HW *et al.* A comparison of the accuracy of ultrasound and computed tomography in common diagnoses causing acute abdominal pain. *Eur. Radiol.* 2011; 21: 1535–45.
- Wu CH, Chen CC, Wang CJ *et al.* Discrimination of gangrenous from uncomplicated acute cholecystitis: accuracy of CT findings. *Abdom. Imaging* 2011; 36: 174–8.
- Chartrand G, Cheng PM, Vorontsov E *et al.* Deep learning: a primer for radiologists. *Radiographics* 2017; 37: 2113–31.
- Soffer S, Ben-Cohen A, Shimon O, Amitai MM, Greenspan H, Klang E. Convolutional neural networks for radiologic images: a radiologist's guide. *Radiology* 2019; 290: 590–606.
- Mentzer RM Jr, Golden GT, Chandler JG, Horsley JS 3rd. A comparative appraisal of emphysematous cholecystitis. *Am. J. Surg.* 1975; 129: 10–5.
- Fry DE, Cox RA, Harbrecht PJ. Gangrene of the gallbladder: a complication of acute cholecystitis. *South. Med. J.* 1981; 74: 666–8.
- Chollet F. Xception: deep learning with depthwise separable convolutions. *arXiv:1610.02357v3*. 2017.
- Russakovsky O, Deng J, Su H *et al.* ImageNet large scale visual recognition challenge. *Int J Comput Vis.* 2015; 115: 211–52.
- Linden A. Measuring diagnostic and predictive accuracy in disease management: an introduction to receiver operating characteristic (ROC) analysis. *J. Eval. Clin. Pract.* 2006; 12: 132–9.
- 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: 837–45.
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159–74.
- Kim T, Choi YH, Choi JH, Lee SH, Lee S, Lee IS. Gallbladder polyp classification in ultrasound images using an ensemble convolutional neural network model. *J. Clin. Med.* 2021; 10: 3585.
- Jeong Y, Kim JH, Chae HD *et al.* Deep learning-based decision support system for the diagnosis of neoplastic gallbladder polyps on ultrasonography: preliminary results. *Sci. Rep.* 2020; 10: 7700.
- Jang SI, Kim YJ, Kim EJ *et al.* Diagnostic performance of endoscopic ultrasound-artificial intelligence using deep learning analysis of gallbladder polypoid lesions. *J. Gastroenterol. Hepatol.* 2021; 36: 3548–55.
- Loukas C, Frountzas M, Schizas D. Patch-based classification of gallbladder wall vascularity from laparoscopic images using deep learning. *Int. J. Comput. Assist. Radiol. Surg.* 2021; 16: 103–13.
- Pang S, Ding T, Qiao S *et al.* A novel YOLOv3-arch model for identifying cholelithiasis and classifying gallstones on CT images. *PLoS One* 2019; 14: e0217647.
- Cheng SM, Ng SP, Shih SL. Hyperdense gallbladder wall sign: an overlooked sign of acute cholecystitis on unenhanced CT examination. *Clin. Imaging* 2004; 28: 128–31.
- Sureka B, Jha S, Rodha MS *et al.* Combined hyperdense gallbladder wall-lumen sign: new computed tomography sign in

acute gangrenous cholecystitis. *Pol J. Radiol.* 2020; 85: e183–7.

- 22 Erlichman DB, Lipman JN, In H, Ye K, Lin J, Goldman I. Diagnostic significance of the CT rim sign in cases of gangrenous cholecystitis. *Clin. Imaging* 2021; 73: 53–6.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Table S1. Interpretation results of the convolutional neural network and the reviewers.

Figure S1. A histogram of the convolutional neural network and the reviewers.

Figure S2. A calibration plot of the convolutional neural network.

Figure S3. Results of the CNN trained on the noncontrast CT and contrast-enhanced CT images used in this study, as well as results of the CNN trained on noncontrast CT images only and contrast-enhanced CT images only.