HOLLOW ORGAN GI

Performance of a machine learning-based decision model to help clinicians decide the extent of lymphadenectomy (D1 vs. D2) in gastric cancer before surgical resection

Chang Liu[1] · Liang Qi[2] · Qiu-Xia Feng[2] · Shu-Wen Sun[2] · Yu-Dong Zhang[2] · Xi-Sheng Liu[2]

Abstract

Background Controversy still exists on the optimal surgical resection for potentially curable gastric cancer (GC). Use of radiologic evaluation and machine learning algorithms might predict extent of lymphadenectomy to limit unnecessary surgical treatment. We purposed to design a machine learning-based clinical decision-support model for predicting extent of lymphadenectomy (D1 vs. D2) in local advanced GC.

Methods Clinicoradiologic features available from routine clinical assignments in 557 patients with GCs were retrospectively interpreted by an expert panel blinded to all histopathologic information. All patients underwent surgery using standard D2 resection. Decision models were developed with a logistic regression (LR), support vector machine (SVM) and auto-encoder (AE) algorithm in 371 training and tested in 186 test data, respectively. The primary end point was to measure diagnostic performance of decision model and a Japanese gastric cancer treatment guideline version 4th (JPN 4th) criteria for discriminate D1 ($pT1 + pN0$) versus D2 ($\geq pT1 + \geq pN1$) lymphadenectomy.

Results The decision model with AE analysis produced highest area under ROC curve (train: 0.965, 95% confidence interval (CI) 0.948–0.978; test: 0.946, 95% CI 0.925–0.978), followed by SVM (train: 0.925, 95% CI 0.902–0.944; test: 0.942, 95% CI 0.922–0.973) and LR (train: 0.886, 95% CI 0.858–0.910; test: 0.891, 95% CI 0.891–0.952). By this improvement, overtreatment was reduced from 21.7% (121/557) by treat-all pattern, to 15.1% (84/557) by JPN 4th criteria, and to 0.7–0.9% (4–5/557) by the new approach.

Conclusions The decision model with machine learning analysis demonstrates high accuracy for identifying patients who are candidates for D1 versus D2 resection. Its approximate 14–20% improvements in overtreatment compared to treat-all pattern and JPN 4th criteria potentially increase the number of patients with local advanced GCs who can safely avoid unnecessary lymphadenectomy.

Keywords Tomography · Gastric cancer · Machine learning · Lymphadenectomy · Decision-support model

---

Dr. CL and Dr. LQ contribute equally to the article in development of methodology, acquisition and analysis of data. They are co-first authors.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00261-019-02098-w) contains supplementary material, which is available to authorized users.

✉ Yu-Dong Zhang
   njmu_zyd@163.com

✉ Xi-Sheng Liu
   njmu_lxs@163.com

Extended author information available on the last page of the article

Introduction

Gastric cancer (GC) is one of most common malignant tumor and the second leading cause of cancer-related death in eastern Asia [1]. Generally, D2 gastrectomy is considered the standard surgical treatment for locally advanced GCs, and the evidence of survival benefit from it was demonstrated on observational studies [2, 3]. However, controversy still exists on the optimal surgical resection for potentially curable GCs because of high complication and death rates following D2 gastrectomy [4, 5]. From recent Japanese gastric cancer treatment guidelines (ver. 4) [6], endoscopic submucosal dissection (ESD), D1 or D1+ gastrectomy is preferred in cT1 and cN0 cancers because of minimal injury,

lower cost, high safety margin, and long-term excellent prognosis. The extent of tumor spread through the gastric wall and nodal status are currently the most important prognostic factors of GCs and make significant influence on selecting the optimal type of surgery. The major challenges exist on correctly defining the tumor invasion (T), and the assessment of N stage is deeply influenced by the extent of lymphadenectomy (D1 vs. D2). As a result, a "stage migration" can occur especially when a limited lymphadenectomy is performed.

The role of computed tomography (CT) in the preoperative staging of GC, even if controversial, may be fundamental for evaluating T and N stage of the gastric disease [7–9]. However, its efficacy in identifying metastatic lymph nodes (LNs) is unsatisfactory, with approximate accuracy of 60% [7, 10–13]. At present, there isn't a uniform standard about the LN metastasis of GCs using either CT or endoscopy [11, 14–16]. The previous researches focused on the size of perigastric LNs, and the size threshold ranges from 5 to 15 mm [8, 11, 17, 18]. However, it is unreliable to evaluate the metastatic LNs by the size of LNs alone, especially when the metastatic LNs were in normal size [15, 19]. Utility of CT may be acceptable for classifying T stage of gastric lesions, especially in locally advanced cases [20]. Evidence shows that the overall accuracy of CT for detecting the depth of tumor invasion ranges from 60 to 90% [8, 21–23]. When using CT, there was a tendency to over stage T2 tumors as T3 tumors; additionally, the detectability of early GC is very low, with a rate of 20% to 50% [24]. Some authors suggested that a combination of signs at CT or a scoring system designed for scaling the risk of tumor invasion and LN metastasis in the general population of patients with GC would be useful [25–29]. However, the CT findings used in previous scoring system were not comprehensive, and some major CT findings of GCs were not investigated. Furthermore, the weights attributed to each of the scoring system variables in these studies were not based on statistical analyses but rather based on subjective assessments. To maximize a future diagnostic scoring system to accurately diagnose tumor invasion and N stage in GCs, the number and type of CT findings need be rigorously determined and clearly evaluated. Machine learning is a newly emerging form of data analysis that uses a series of data mining algorithms or statistical tools analyses of high-throughput features to obtain predictive or prognostic information. By building appropriate models with refined features, it has shown successful assessment and prediction abilities in various challenging clinical tasks [30–32].

Therefore, our purpose is to design a machine learning-based decision-support model that relies primarily on the evidence-based clinical and CT imaging data, with the aim to elaborate on this information for preoperatively predicting extent of lymphadenectomy (D1 vs. D2) in patients with GC.

Materials and methods

Patients

This was a retrospective study involving routine at a single medical center. Ethics Committee approval was granted by Local Institutional Ethics Review Board with a waiver of written informed consent. All procedures performed in studies involving human participants were in accordance with the 1964 Helsinki Declaration and its later amendments.

We queried our institution's pathology database to derive all histologically proved cases of GCs between January 2014 and December 2016. Total 789 histologically proved cases of GCs were identified in this query. Among all patients, we included patients who fulfilled the following criteria: (1) patients who had preoperative CT images with optimal gastric distension, (2) patients whose primary gastric lesion were detectable on axial scans of CT images, and (3) patients who underwent standard D2 gastrectomy [6] and without the history of preoperative chemoradiation therapy. According to the initial inclusion criteria, 232 patients were excluded because some patients were with history of preoperative chemoradiation therapy ($n = 20$) or with an endoscopic resection of the cancer ($n = 131$), some were with end-stage GCs failed to undergo radical gastrectomy ($n = 36$), others were with lack of preoperative CT imaging or underwent CT examination from outside institutions ($n = 45$). At last, 557 patients with histopathologic-confirmed GCs who underwent gastrointestinal CT scan before standard D2 gastrectomy were retrospectively included. Median time interval between CT examination and surgery was 9 days (range 6–14 days).

CT image acquisition

All gastric CT studies were performed with a 64-slice scanner (SOMATOM, Definition AS+, Siemens, Forchheim, Germany). The CT scans, covering from the liver to hypogastric region, were acquired during a breath-hold with the patient supine. All patients drank 600–800 mL of water before CT examination. The CT scan parameters were as follows: 120 kVp, effective 160 mAs, rotation time 0.5 s, detector collimation 32×1.2 mm, field of view 350×350 mm, matrix 512×512, and reconstruction section thickness 1.5 mm. Patients were injected with non-ionic contrast material (Ultravist; 300 mg I/mL, Bayer Schering Pharma AG, Berlin, Germany) with antecubital venous access at a rate of 3.0 mL/s, a total of 90–120 mL (1.5 mL/kg of body weight) was injected by using the CT-compatible power injector during arterial and

parenchymal phase. The scanning delay for arterial imaging was determined by using automated scan-triggering software (Care-Bolus; Siemens Medical Systems, Iselin, NJ). Arterial scanning automatically began 7.0 s after the trigger attenuation threshold (100 HU) was reached at the level of the supraceliac abdominal aorta. At a delay of 40 s after arterial scanning, parenchymal scanning began.

Imaging analysis

All CT images were retrospectively interpreted by two radiologists (C.L. and L.Q.) who had 3-year (C.L.) and 10-year (L.Q.) experience in reading gastrointestinal images. Both the readers were members of the institution's GI disease management team and had read > 500 GI CT studies. Before the study, they were provided with a whole day lecture-based and hands-on instruction that explained in detail a structured gastrointestinal image reporting scheme, in which 40 cases randomly selected from institutional RIS/PACS database (not involved in this study cohort) were reviewed individually, scored, and then reviewed as a group.

Our gastrointestinal imaging scoring scheme, referring to the methods of previous studies [7, 8, 16, 33, 34], included several structured imaging features of GCs: (1) region of tumor involvement (1, cardia; 2, fundus; 3, body; 4, antrum), (2) maximum length of the tumor (L_{max}), (3) tumor margin (0, sharply defined; 1, ill-defined), (4) tumor growth pattern (0, intra-gastric; 1, mixed), (5) intra-tumor necrosis (0, absent; 1, present), (6) tumor ulceration (0, absent; 1, present), (7) serosal invasion (0, absent; 1, present), wherein, the serosal invasion was defined as an irregular or nodular outer margin of the serosal layer and/or a dense band-like perigastric fat infiltration [13], (8) perigastric fat invasion (0, absent; 1, present), (9) radiologic T stage (T1–4), which referred to the standards of Kim et al. and Ahn et al. [7, 8], (10) peritoneal seeding sign (0, absent; 1, present), (11) tumor arterial enhancement (TAE), which measured the CT HU at a represent region of interest (ROI), (12) tumor parenchymal enhancement (TPE), (13) the tumor contrast between arterial and parenchymal enhancement (CAP), and (14) the short axis length of the largest LN discernible (LND) was individually determined, and LNDs of less than 5 mm were rounded down to 0 mm in this study. The two radiologists independently reviewed the gastric images. Any disagreement in imaging interpretations between the readers were discussed until a final standard consensus was generated. The measurable radiologic features such as L_{max}, LND, TAE, TPE, and CAP were averaged between the two radiologists. This procedure was performed after an inter-observer reliability analysis (Supplemental Data, S1). Representative imaging examples (Fig. 1) are provided to describe applied methodologies for image interpretation along with different radiologic categories.

Histopathology and reference standard

All patients underwent standard D2 gastrectomy. Both N1 and N2 LNs were systematically dissected, regardless of their CT images. N3 and N4 LNs were dissected if they were visible and palpable at surgery. Node packets were sent separately and submitted for microscopic examination to evaluate for the presence of metastasis. The pT and pN stage were reported according to the 7th American Joint Committee on Cancer Staging (AJCC) system [20]. Regarding the reference standard, patients with pT1 and
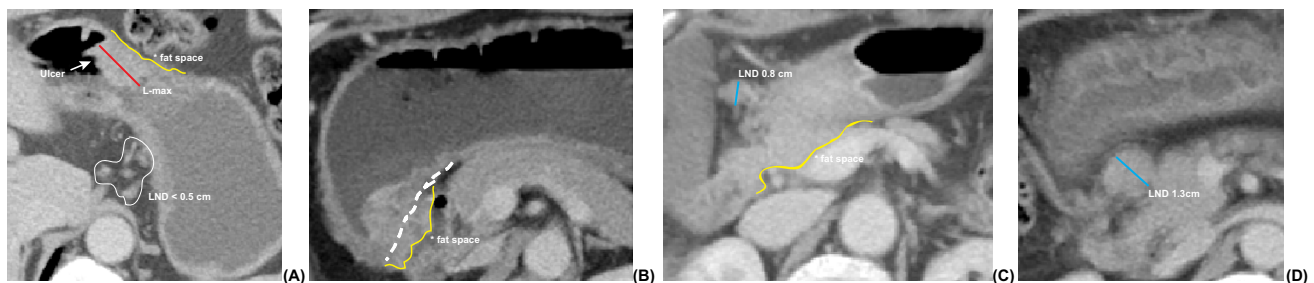

Fig. 1 The illustration of typical radiologic features of gastric cancer. **a** A 68-year-old man presented with an intra-cavity mass in anterior wall of body of stomach; the L_{max} of the mass is shown with red line; a small ulcer is determined at internal surface of the mass (arrow); the adjacent serosal surface is smooth, implying a T2-stage lesion; and perigastric fat space is clear (*, yellow outline), suggesting a negative fat invasion; clustering lymph nodes (LNs) are detected in N1 station with size of < 0.5 cm, suggesting a negative LN finding. **b** A 63-year-old man was found with an irregular mass in antrum of stomach, which presented with undefined boundary (yellow outline) and blurry perigastric fat space (*), suggesting an extra-stomach extension and perigastric fat infiltration (T4a stage). **c** A 69-year-old man was found with a solid mass in body of stomach. The mass had an extra-gastric wall extension and the perigastric fat space is disappeared (*), implying a suspected adjacent organ involvement (T4b stage). The measured LN size (LND) in this case is less than 1 cm (blue line), demonstrating negative LN finding. **d** A mass in antrum of stomach presented with an enlarged LN (blue line, LND 1.3 cm), demonstrating a typical metastatic LN

pN0 stage were classified into D1-resection group, and patients with ≥ pT1 and ≥ pN1 stage were classified into D2-resection group.

Model development

The patients were divided into two groups (2:1) in a randomized fashion: one group was designated train group ($n = 371$) for training the decision model, and the other group was designated test group ($n = 186$) for the evaluation of the accuracy of decision model. This randomized fashion for the classification of patients is an advisable way to avoid over fitting. Clinicoradiologic candidate predictors such as age, sex, tumor location, L_{max}, tumor margin, tumor growth pattern, intratumoral necrosis, intratumoral ulceration, serosal invasion, perigastric fat invasion, peritoneal seeding sign, radiologic T stage, TAE, TPE, CAP, and LND were included as input variables. Before the machine learning analysis, continuous variables such as L_{max}, LND, TAE, TPE, and CAP were binarily reclassified using a receiver operating characteristic (ROC) analysis (Supplemental Data, S2). Values lower than the defined ROC threshold were designated score 0, and those higher the threshold were designated score 1.

The decision models, discriminating patients who were candidates for D1 versus D2 resection, were developed using a logistic regression (LR) classification, a support vector machine (SVM) and an auto-encoder (AE) machine learning analysis, respectively. In LR analysis, variables associated with D2-endpoint by univariate analyses at a significant level were candidates for stepwise multivariate analysis. Predictive model was formulated based on the results of multivariate regression analysis. The details of SVM and AE networks regarding feature selection, hyper-parameter optimization and model development are explained in Supplemental Data, S3. At the end of classification, the output of classifiers was converted to a probability (PI). The value of PI indicates individual surgical index for D2 resection.

As primary objective of the study is to investigate potential benefit of new decision model for preoperatively discriminating between D1 and D2 resection, we used Japanese gastric cancer treatment guideline version 4th (JPN 4th) as a control method for comparison [6]. In JPN 4th criteria, patients were classified into cD1 group if a cT1 and cN0 stage was determined at CT imaging, and patients were classified into cD2 group if $a ≥ cT1$ and $≥ cN1$ stage was determined. The threshold for cN1 at CT imaging was set at LND > 10 mm. Nodes were considered positive or negative on basis of the groups as a whole; that is, positive if one or more LNs metastases were discernible in the group and negative if none was discernible.

Statistical analysis

Performance analysis comprised three activities. First, model discrimination was quantified with a ROC curve analysis. The sensitivity and specificity were computed for the cutoff on the ROC curve that maximizes the product of both measures. Second, a decision curve analysis, proposed by Vickers and Elkin [35], was performed to evaluate the clinical usefulness between JPN 4th and decision models. The decision curve estimates the net benefit of a model by the difference between the number of true-positive and false-positive rates, weighted by the odds of the selected threshold probability of risk. The statistical analysis was conducted with an R-package (version 3.3.4; http://www.Rproject.org). The reported statistical significance levels were all two-sided, with statistical significance set at 0.05.

Results

Of 557 patients included, pD1 was diagnosed in explanted tissue of 121/557 (21.7%) patients. The mean age was 60.8 years (standard deviation 10.8 years). The distribution of pathologic tumor stage was pT1 27.6% (154/557), pT2 11.3% (63/557), pT3 19.7% (110/557), and pT4 41.3% (230/557). The distribution of pathologic nodal status was pN0 40.0% (223/557), pN1 14.5% (81/557), pN2 15.3% (85/557), pN3a 17.4% (97/557), and pN3b 12.7% (71/557). The detail histopathologic results of nodal packages resected at surgery were summarized in Table 1.

Radiologic finings of GCs in 557 patients examined were listed in Table 2. In SVM analysis, among all 27 clinicoradiologic features, feature of age, sex, intramural growth, and intratumoral necrosis tested with false-discovery-rate U-test ($p > 0.05$) were initially excluded. SVM with recursive feature elimination (RFE) analysis selected total 11 most important features which produced largest AUC and smallest bias over the bootstrapping validation (Fig. 2). In LR model, eight independent features were determined at significant level with the multivariate regression analysis. The coefficients of independent features and adjusted odds ratio (OR) were summarized and compared to SVM weights in Table 3. The distribution of decision scores obtained from LR, SVM and AE were illustrated in Fig. 3. Significant differences of decision scores were evidenced between pD1 and pD2 group (LR: 0.50 ± 0.16 vs. 0.82 ± 0.15, $p < 0.001$; SVM: -0.18 ± 0.26 vs. 0.96 ± 0.38, $p < 0.001$; AE: 0.21 ± 0.17 vs. 0.95 ± 0.18, $p < 0.001$).

The ROC analysis shows that AE model yielded the highest AUC (training: AE vs. SVM, $p < 0.001$; AE vs. LR, $p < 0.001$; test: AE vs. SVM, $p > 0.05$; AE vs. LR, $p < 0.001$), followed by SVM model (training: SVM vs. LR, $p < 0.001$; test: SVM vs. LR, $p < 0.001$) and LR model, respectively

Table 1 Histopathologic results of nodes resected at surgery

Packages	No. of nodes resected	No. of positive nodes	Percentage of positive nodes
N1	1625	245	15.1
N2	574	113	19.7
N3	6254	1359	21.7
N4	3607	661	18.3
N5	757	117	16.5
N6	1411	239	16.7
N7	911	136	14.9
N8	847	108	12.8
N9	712	109	15.3
N10	193	24	12.4
N11	404	77	19.1
N12	328	30	9.1
N13	62	14	22.6
N14	88	7	7.9
N15	13	0	0
N16	35	2	5.7
Nodal status			
N0	6817	0	0
N1	2689	119	4.4
N2	2671	383	14.3
N3a	3191	1046	32.8
N3b	2852	1776	62.3
Total	18,219	3305	18.1

No. regional lymph node metastasis, *N1* 1 or 2 positive lymph nodes, *N2* 3 to 6 positive lymph nodes, *N3a* 7 or more positive lymph nodes, *N3b* more than 15 positive lymph nodes

(Fig. 4). Using an optimal threshold (LR score > 0.73; SVM score > 0.71; AE score > 0.99) that maximize the Youden Index, AE model achieved best diagnostic performance regarding both high sensitivity and specificity, followed by SVM and LR model. Compared to JPN 4th criteria, three models had significant improvement in diagnostic specificity (Table 4).

As part of this study, we considered the implications of using decision models as a triage test to decide which men should undergo D1 resection, purposing to avoid unnecessary lymphadenectomy. In the worst-case scenario in our medical center, all the men with histopathology-confirmed GCs would undergo D2 resection. This procedure resulted in 21.7% (121/557) overtreatment using pathologic stage as reference standard. If JPN 4th criteria was used, 15.1% (84/557) patients would be over treated by active D2 resection and 2.7% (15/557) patients would be undertreated by D1 resection. When a decision-support model was used, the false-positive rate was reduced to 0.7% (4/557) by LR model, to 0.9% (5/557) by SVM model and to 0.9% (5/557) by AE model. While the false negative rate was LR 17% (95/557), SVM 10.2% (57/557), and AE 8.8% (49/557). If

Table 2 Radiologic results of GC patients between pathologic D1 and D2 group

Imaging findings	pD1 (*n* = 121)	pD2 (*n* = 436)	*p*
Origination involved			
Cardia	21/121 (17.4)	163/436 (37.4)	< 0.001
Fundus	2/121 (1.7)	70/436 (16.1)	< 0.001
Body	59/121 (48.8)	300/436 (68.8)	< 0.001
Antrum	83/121 (68.6)	224/436 (51.4)	0.001
L_{max} (cm)[b]	2.7 ± 1.1	5.2 ± 2.1	< 0.001
Tumor margin (ill-defined)	47/121 (38.8)	310/436 (71.1)	< 0.001
Tumor growth pattern			
Intramural	121/121 (100)	434/436 (99.5)	0.455
Extramural	0/121 (0)	66/436 (15.1)	< 0.001
Transmural	0/121 (0)	65/436 (14.9)	< 0.001
Diffuse	1/121 (0.8)	102/436 (23.4)	< 0.001
Invasive	2/121 (1.7)	196/436 (45.0)	< 0.001
Intratumoral necrosis	2/121 (1.7)	9/436 (2.1)	0.774
Intratumoral ulcer	49/121 (40.5)	317/436 (72.7)	< 0.001
Serosal invasion	0/121 (0)	38/436 (8.7)	0.001
Blurry fat space	3/121 (2.5)	122/436 (28.0)	< 0.001
Peritoneal seeding sign	0/121 (0)	38/436 (6.8)	0.001
Radiologic T stage[a]			
T1	38/121 (31.4)	15/436 (3.4)	< 0.001
T2	55/121 (45.5)	69/436 (15.8)	< 0.001
T3	28/121 (23.1)	180/436 (41.3)	< 0.001
T4	0/121 (0)	172/436 (39.4)	< 0.001
Radiologic N status[a]			
N1 (stations 1–6)	43/121 (35.5)	365/436 (83.7)	< 0.001
N2 (stations 7–11)	13/121 (10.7)	188/436 (43.1)	< 0.001
N3 (stations 12–14)	1/121 (0.8)	63/436 (14.4)	< 0.001
N4 (stations 15, 16)	2/121 (1.7)	61/436 (14.0)	< 0.001
LND (cm)[b]	0.21 ± 0.36	0.73 ± 0.65	< 0.001
TAE[b]	1.03 ± 0.27	1.14 ± 0.26	< 0.001
TPE[b]	1.09 ± 0.26	1.25 ± 0.27	< 0.001
CAP[b]	0.94 ± 0.17	1.02 ± 0.33	< 0.001

Values are no. of findings, and values in parentheses are percentages unless indicated otherwise. Continuous data are analyzed by *t* test, counted data are analyzed by χ^2 tests

L_{max} maximum length of the tumor, *TAE* arterial enhancement of the tumor, *TPE* parenchymal enhancement of the tumor, *CAP* tumor contrast between arterial and parenchymal enhancement, *LND* the short axis diameter of the largest lymph nodes discernible

[a]Radiologic T and N status are based on Japanese Research Society for Gastric Cancer Guideline

[b]Reports are means and standard deviation

$a < 10\%$ false-judge rate is clinically tolerated, AE model is the only one who satisfies this criteria. Considering high risk of false-positive cases produced by JPN 4th criteria, we tested the incremental value of decision-support score for guiding decision-making with a decision curve analysis
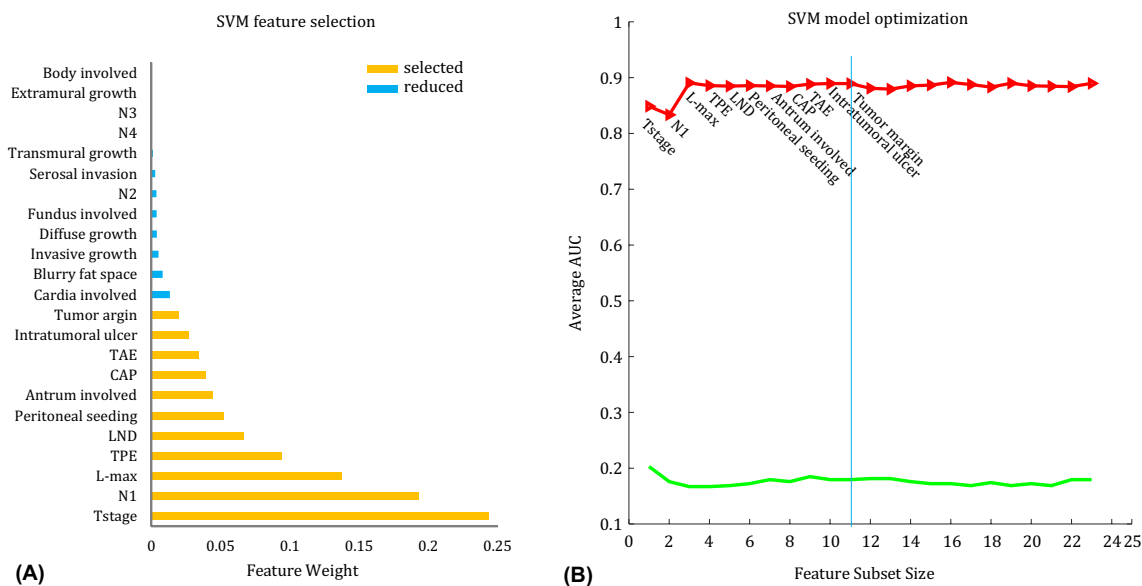
Fig. 2 Results of feature selection, feature ranking, and model construction with RFE-SVM analysis. **a** Distribution of weight for Features with RFE analysis. **b** SVM classifier is trained by adding ranked feature one-by-one. The iteration repeated until the desired number of features was reached. The desired features and classifier parameters are selected to be those resulting in best accuracy and least estimated bias over cross-validation (blue line)

Table 3 Features weights and coefficients estimated by SVM and LR analysis

Feature	SVM weight[b]	LR coefficient[b]	OR	95% CI	
				Low	Upper
T stage	0.244	0.781	2.184	1.457	3.274
N1	0.193	0.705	2.024	1.141	3.59
L_{max} (≤ 3.7 cm vs. > 3.7 cm)[a]	0.137	1.31	3.705	1.888	7.268
TPE (≤ 1.11 vs. > 1.11)[a]	0.094	0.568	1.765	1.036	3.008
LND (≤ 0.3 cm vs. > 0.3 cm)[a]	0.067				
Peritoneal seeding sign	0.052				
Antrum involved	0.045				
CAP (≤ 1.01 vs. > 1.01)[a]	0.039	0.584	1.794	0.975	3.299
TAE (≤ 0.99 vs. > 0.99)[a]	0.034				
Intratumoral ulcer	0.027				
Tumor margin	0.020				
Cardia involved		0.688	1.990	1.023	3.871
Blurry fat space		1.181	3.257	0.872	12.164
Invasive growth		1.544	4.685	1.012	21.701

OR odds ratio with LR analysis, *CI* confidence interval

[a]Threshold obtained from a ROC analysis in Supplemental Data S1

[b]Only the features selected into the training model are listed

(Fig. 5). Decision curves were estimated using JPN 4th and the decision score produced by AE model. It shows that adding decision score to JPN 4th criteria produces significant benefit than use of JPN 4th criteria solely.

Discussion

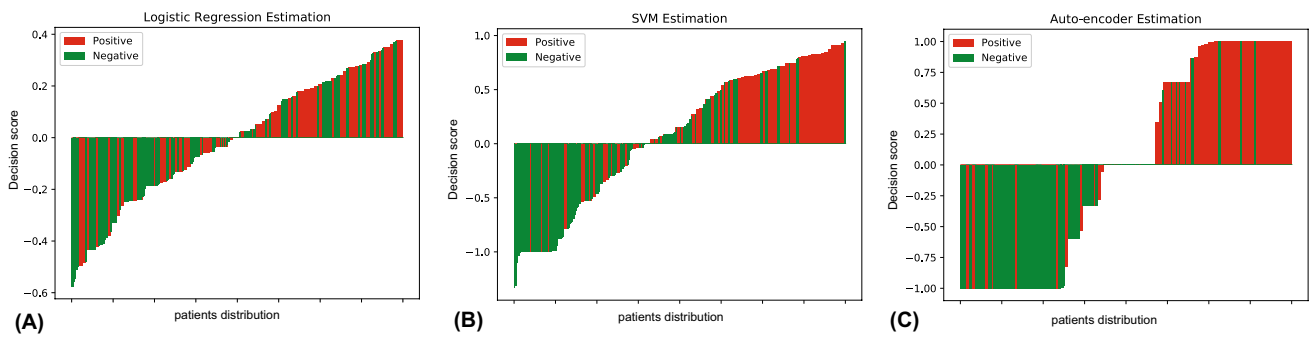The results of our study draw a conclusion that the decision model derived from machine learning analysis

Fig. 3 Distribution of decision score obtained by LR (**a**), SVM (**b**) and AE (**c**) analysis
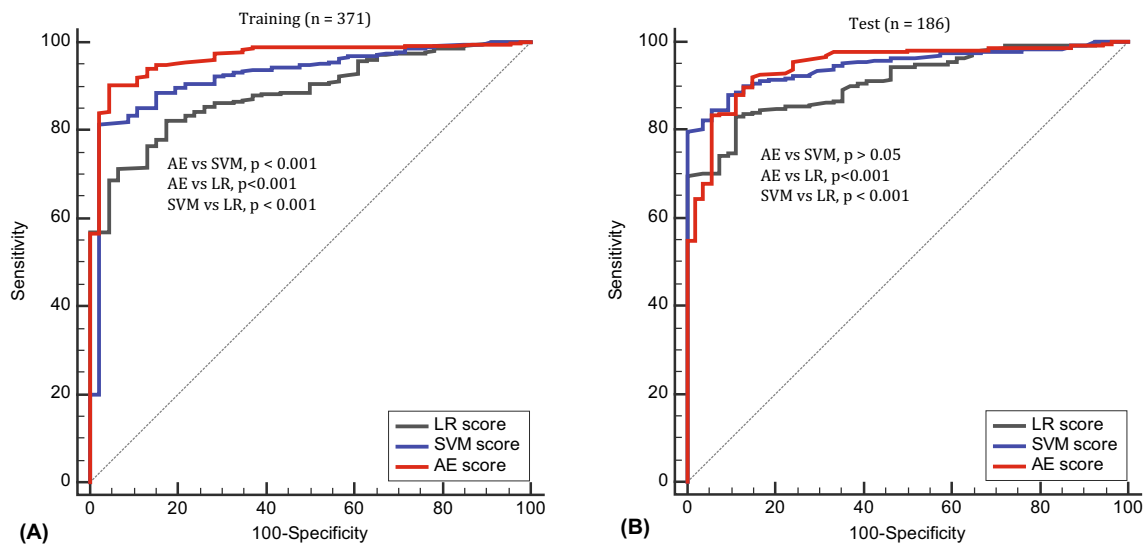

Fig. 4 ROC analysis of decision models with LR, SVM, and AE algorithms

Table 4 The diagnostic performance of decision score derived from LR, SVM, and AE analysis, and compared to JPN 4th criteria

Group	AUC (95% CI)	SEN	SPE	PPV	NPV	Cutoff
Training (*n* = 371)						
LR	0.886 (0.858–0.910)	82.2	78.3	98.0	25.7	> 0.73
SVM	0.925 (0.902–0.944)	89.9	78.3	98.1	37.9	> 0.71
AE	0.965 (0.948–0.978)	90.2	89.1	99.1	41.8	> 0.99
Test (*n* = 186)						
LR	0.891 (0.891–0.952)	84.9	77.8	96.6	23.4	> 0.73
SVM	0.942 (0.922–0.973)	87.2	87.4	97.2	39.3	> 0.71
AE	0.946 (0.925–0.978)	89.6	87.0	98.8	39.7	> 0.99
JPN 4th (overall patients)	–	96.6	30.6	83.4	71.2	cT1 + cN0

SEN sensitivity, *SPE* specificity, *PPV* positive predictive value, *NPV* negative predictive value, *CI* confidence interval

achieved good operation performance in local advanced GCs. Songun et al. indicated that D2 lymphadenectomy is associated with lower locoregional recurrence and GC-related death rates than D1 surgery. However, D2 patients were also associated with significantly higher operative mortality, higher complication rate and higher
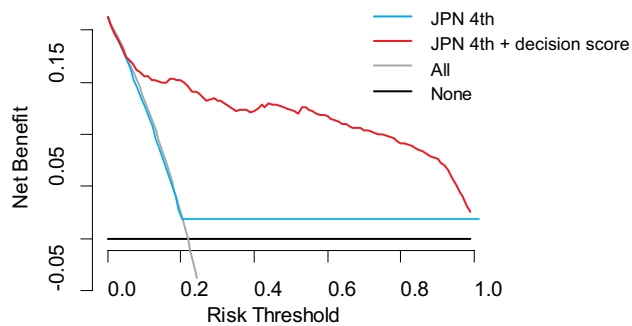
Fig. 5 Decision curve analysis of decision score for discriminating D1 versus D2 lymphadenectomy. The *y*-axis measures the net benefits, and the *x*-axis is the risk threshold. The black line represents the assumption that no patients undergo D2 lymphadenectomy, the gray line represents the assumption that all patients have D2 lymphadenectomy. The blue line represents the results estimated by JPN 4th criteria. The red line represents the results estimated by a combination of decision-support model with JPN 4th. It shows that, the JPN 4th criteria has no benefit compared to treat-all and treat-none pattern, while if decision score is added, it results in more benefits for patients' discrimination

reoperation rate than D1 patients [4]. In addition, ESD, D1 or D1+ gastrectomy is preferred in cT1 and cN0 cancers because of minimal injury, lower cost, high safety margin, and long-term excellent prognosis [6].Due to the extremely important significance in guidance of preoperative therapeutic schedules and selection of extent of lymphadenectomy, this decision-support calculator with high accuracy and good discrimination is in imminent need. The system is superior to JPN 4th classification regarding high specificity and low false-positive prediction rate. This would connect to better selection of optimal type of treatment, thus allows to minimal injury, high safety margin, and long-term excellent prognosis. Therefore, we believe our research findings can play an important role and make a significant difference in current clinical work of local advanced GCs.

As inhomogeneous biologic feature of GCs, the treatment and prognosis of GCs is significantly different among histological type, tumor invasion depth, and adjacent organ morphologic relations [36–40]. Surgical resectability of GCs is largely dependent on the preoperative evaluation of the features such as tumor invasion depth, N status, adjacent organ involvement, peritoneal carcinomatosis, and distant metastases, which can be determined at imaging-based findings. However, there is still a matter of extensive debate on "stage migration" caused by conventional staging scheme that relies on morphologic features at CT images. This contributes to significant increase in over diagnosis of early disease and under-diagnosis of advance cancers. Generally, the CT finding most suggestive of metastatic LNs involvement is nodal size of 1 cm or greater on CT images, presenting with

the diagnostic accuracies of 50–80% [41, 42]. In current study, we define LNs as metastatic when the short axis diameter (LND) is over 1 cm on CT images. Of the total 557 dissected LN diseases, 455 patients with nodes were less than 1 cm in size and 258/455 (56.7%) of them were metastatic histologically; and of 102 dissected LNs larger than or equal to 1 cm in size, benign LNs were confirmed histologically in 26/102 cases (25.4%). The limitation of size criteria for determination of N status has long been recognized, as the inhomogeneous biologic feature of GCs, approximately 40% of metastatic LNs occur in the body will have a 'normal' size. Additionally, "Stage migration" might be notified in classifying tumor invasion depth. In current study, we used a multi-detector CT including conventional 2D axial, coronal or sagittal multiplanar reformatted images for the evaluation of T stage. Of 557 resected lesions, the accuracy of CT for diagnosing pT1, pT2, pT3, and pT4 is 45/154 (29.2%), 25/63 (39.7%), 47/110 (42.7%), and 119/230 (51.7%), respectively, significant difference with those reported previously [8, 21–24]. Stage migration might be associated with variabilities in image scanning, reconstruction and individual experience of radiologist. Using this staging strategy, i.e., JPN 4th criteria, for guiding surgical option resulted in 84/557 (15.1%) overtreatment of patients who were true candidates for D1 resection.

Decision-support model brought us new insights to advance diagnosis and surgical procedure. Biological characters of GCs such as histological type, tumor invasion depth, and adjacent organ morphologic relations could be reflected and predicted by deep analysis of large scale imaging features. This data mining approach might be helpful for improvement of clinical decision. Hong et al. developed a scoring system to predict the risk of No. 10 LN metastasis for advanced upper GCs, and the AUC of this scoring system was 0.82 [25]. In current study, we developed an image reporting scheme for introducing CT features of primary tumors and perigastric LNs. All variables reported were highly standardized and incorporated into the decision model using two novel machine learning methods and compared them to a conventional LR analysis. Use of those two machine learning approaches, advantage in discriminating D1 versus D2 resection was clearly observed as compared with conventional LR approach and two-point JPN 4th criteria. The diagnostic specificity was improved and overestimate was reduced when new scoring scheme was used.

Even inconsistent results were observed between SVM and LR for the feature selection, both of them demonstrated that higher radiologic T stage, positive findings at N1 station, larger L_{max}, and higher TPE were significant predictors for discriminating between D1 and D2 resection. Our findings were comparable to those of previous studies, which reported that LN metastasis strongly correlates to tumor size (L_{max} in this study) and tumor depth of invasion (radiologic

T stage in this study) [25, 43, 44]. Additionally, Miles et al. found that tumor contrast-enhanced density was strongly associated with tumor blood supply, which made it easier having distal invasion [45]. Fukuya et al. reported that the larger short axis diameter of LN (LND in this study) was the more probability of LN metastasis [46]. Thus, we suggested that GC is a rich-blood supplied malignancy; the size and invasion depth of the tumor increase consistently with tumor progression; and eventually, tumor penetrates through the muscle layer and infiltrates the serosa layer, which contains abundant capillary lymphatic net, the incidence of distant invasion and metastasis thus increase. We found AE model gave the best results against SVM and LR. Unlike conventional machine learning classifiers, the AE deep learning network allows to learn good features automatically from data, and thus avoids manually-designed feature extractors. Secondly, AE imposes sparsity on the mapped features (i.e., responses of hidden nodes), thus avoiding the problem of trivial solutions when the dimensionality of hidden features is more than that of the input features. Besides, we used unsupervised initialization in the pre-training stage, which prevents the later supervised training from falling into the bad local minimum. Therefore, we support that the AE deep learning is a new way and deserved to be used in current or the future study.

The limitations of our study include the lack external validation for the new decision-support system. Our results were based on clinical data obtained from an eastern country at a single institution and may not be generalized to other populations of patients with GCs. Eastern countries have higher GC morbidity and more advanced GC patients than western countries. This bias in population distribution leads to lower early GC patients in this study, which might make influence on the statistical analysis. Multicenter validation with a larger sample size is mandatory to acquire high-level evidence for clinical application. In addition, Radiomics is a rising field of quantitative imaging that capture tumor phenotype characteristics non-invasively using advanced imaging features. Radiomics may help to determine relationships between such features and the underlying pathophysiology; therefore, a combination of Radiomics signature may improve the ability to predict the stage of GC in further studies.

Conclusions

In current study, we designed a decision-support model that relies primarily on the evidence-based clinicoradiologic data, aiming for preoperatively predicting extent of lymphadenectomy in patients with GC. Using a machine learning-based analysis for data mining and incorporating,

we suggested that application of such model-dependent approach offers an alternative to determine tumor biological characters and predict clinical status of the patient, and thus permit better-informed pretreatment decisions.

Compliance with ethical standards

Conflict of interest The authors have nothing to declare.

Ethical approval The study was approved by the Institutional Review Board of the First Affiliated Hospital of Nanjing Medical University.

Informed consent Written informed consent was waived by the Institutional Review Board.

References

1. Siegel RL, Miller KD, Jemal A (2018) Cancer statistics, 2018. *CA: a cancer journal for clinicians*; **68**(1): 7–30.
2. Wu CW, Hsiung CA, Lo SS et al (2006) Nodal dissection for patients with gastric cancer: a randomised controlled trial. The Lancet Oncol **7**(4): 309–315.
3. Sasako M, Sano T, Yamamoto S et al (2008) D2 lymphadenectomy alone or with para-aortic nodal dissection for gastric cancer. N. Engl. J. Med **359**(5): 453–462.
4. Songun I, Putter H, Kranenbarg EM et al (2010) Surgical treatment of gastric cancer: 15-year follow-up results of the randomised nationwide Dutch D1D2 trial. The Lancet Oncol **11**(5): 439–449.
5. Cuschieri A, Weeden S, Fielding J et al (1999) Patient survival after D1 and D2 resections for gastric cancer: long-term results of the MRC randomized surgical trial. Surgical Co-operative Group. Br. J. Cancer **79**(9–10): 1522–1530.
6. Japanese Gastric Cancer A. Japanese gastric cancer treatment guidelines 2014 (ver. 4). *Gastric cancer: official journal of the International Gastric Cancer Association and the Japanese Gastric Cancer Association* 2017; **20**(1): 1–19.
7. Ahn HS, Kim SH, Kodera Y, Yang HK (2013) Gastric cancer staging with radiologic imaging modalities and UICC staging system. Dig. Surg **30**(2): 142–9.
8. Kim JW, Shin SS, Heo SH et al (2012) Diagnostic performance of 64-section CT using CT gastrography in preoperative T staging of gastric cancer according to 7th edition of AJCC cancer staging manual. Eur. Radiol **22**(3): 654–662.
9. Lee SL, Ku YM, Jeon HM, Lee HH (2017) Impact of the Cross-Sectional Location of Multidetector Computed Tomography Scans on Prediction of Serosal Exposure in Patients with Advanced Gastric Cancer. Ann Surg Oncol **24**(4): 1003–1009.
10. Hur J, Park MS, Lee JH et al (2006) Diagnostic accuracy of multidetector row computed tomography in T- and N staging of gastric cancer with histopathologic correlation. J Comput Assist Tomo **30**(3): 372–377.
11. Kim SH, Kim JJ, Lee JS et al (2013) Preoperative N staging of gastric cancer by stomach protocol computed tomography. J Gastric Cancer **13**(3): 149–156.

12. Stabile Ianora AA, Telegrafo M, Lucarelli NM et al(2017) Comparison between CT Net enhancement and PET/CT SUV for N staging of gastric cancer: A case series. Ann Med Surg **21**: 1–6.

13. Kim HJ, Kim AY, Oh ST et al (2005) Gastric cancer staging at multi-detector row CT gastrography: comparison of transverse and volumetric CT scanning. Radiology **236**(3): 879–885.

14. Maccioni F, Marcelli G, Al Ansari N et al (2010) Preoperative T and N staging of gastric cancer: magnetic resonance imaging (MRI) versus multi detector computed tomography (MDCT). Clin Ter **161**(2): e57–62.

15. Huang CM, Xu M, Wang JB et al (2014) Is tumor size a predictor of preoperative N staging in T2-T4a stage advanced gastric cancer? Surg Oncol **23**(1): 5–10.

16. Hwang SH, Kim HI, Song JS, Lee MH, Kwon SJ, Kim MG (2016) The Ratio-Based N Staging System Can More Accurately Reflect the Prognosis of T4 Gastric Cancer Patients with D2 Lymphadenectomy Compared with the 7th American Joint Committee on Cancer/Union for International Cancer Control Staging System. J Gastric Cancer **16**(4): 207–214.

17. Dorfman RE, Alpern MB, Gross BH, Sandler MA (1991). Upper abdominal lymph nodes: criteria for normal size determined with CT. Radiology **180**(2): 319–322.

18. Chen CY, Hsu JS, Wu DC et al (2007) Gastric cancer: preoperative local staging with 3D multi-detector row CT–correlation with surgical and histopathologic results. Radiology **242**(2): 472–482.

19. Peng CW, Wang LW, Zeng WJ, Yang XJ, Li Y (2013) Evaluation of the staging systems for gastric cancer. J Surg Oncol **108**(2): 93–105.

20. Washington K (2010) 7th edition of the AJCC cancer staging manual: stomach. Ann Surg Oncol **17**(12): 3077–3079.

21. Kumano S, Okada M, Shimono T et al (2012) T-staging of gastric cancer of air-filling multidetector-row CT: comparison with hydro-multidetector-row CT. Eur J Radiol **81**(11): 2953–2960.

22. Lee IJ, Lee JM, Kim SH et al (2009) Helical CT evaluation of the preoperative staging of gastric cancer in the remnant stomach. AJR Am J Roentgenol **192**(4): 902–908.

23. Kumano S, Murakami T, Kim T et al (2005) T staging of gastric cancer: role of multi-detector row CT. Radiology **237**(3): 961–966.

24. Fairweather M, Jajoo K, Sainani N, Bertagnolli MM, Wang J (2015) Accuracy of EUS and CT imaging in preoperative gastric cancer staging. J Surg Oncol **111**(8): 1016–1020.

25. Hong ZL, Chen QY, Zheng CH et al (2017) A preoperative scoring system to predict the risk of No. 10 lymph node metastasis for advanced upper gastric cancer: a large case report based on a single-center study. Oncotarget **8**(45): 80050–80060.

26. Xu S, Feng L, Chen Y et al (2017) Consistency mapping of 16 lymph node stations in gastric cancer by CT-based vessel-guided delineation of 255 patients. Oncotarget **8**(25): 41465–41473.

27. Peng C, Liu J, Yang G, Li Y (2018) The tumor-stromal ratio as a strong prognosticator for advanced gastric cancer patients: proposal of a new TSNM staging system. J Gastroenterol **53**(5):606–617

28. Choi YY, An JY, Katai H et al (2016) A Lymph Node Staging System for Gastric Cancer: A Hybrid Type Based on Topographic and Numeric Systems. PloS one **11**(3): e0149555.

29. Chen J, Chen C, He Y, Wu K, Wu H, Cai S (2014) A new pN staging system based on both the number and anatomic location of metastatic lymph nodes in gastric cancer. J Gastrointest Surg **18**(12): 2080–2088.

30. Aerts HJ, Velazquez ER, Leijenaar RT et al (2014) Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Na Commun **5**: 4006.

31. Kermany DS, Goldbaum M, Cai W et al (2018) Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Cell **172**(5): 1122–31 e9.

32. Huang YQ, Liang CH, He L et al (2016) Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. J Clin Oncol **34**(18): 2157–2164.

33. Kim AY, Kim HJ, Ha HK (2005) Gastric cancer by multidetector row CT: preoperative staging. Abdom Imaging **30**(4): 465–472.

34. Makino T, Fujiwara Y, Takiguchi S et al (2011) Preoperative T staging of gastric cancer by multi-detector row computed tomography. Surgery **149**(5): 672–679.

35. Vickers AJ, Elkin EB (2006) Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making **26**(6): 565–574.

36. Zhang CD, Ning FL, Zeng XT, Dai DQ (2018) Lymphovascular invasion as a predictor for lymph node metastasis and a prognostic factor in gastric cancer patients under 70 years of age: A retrospective analysis. Int J Surg **53**:214–220

37. Ozawa Y, Chiba N, Hikita K et al (2017) Long-Term Survival of a Gastric Neuroendocrine Carcinoma Patient with Extra-Regional Lymph Node Metastases. Gan To Kagaku Ryoho **44**(4): 333–336.

38. Naffouje SA, Salti GI (2017) Extensive Lymph Node Dissection Improves Survival among American Patients with Gastric Adenocarcinoma Treated Surgically: Analysis of the National Cancer Database. J Gastric Cancer **17**(4): 319–330.

39. Marin Cordova NE, Yan-Quiroz EF, Diaz Plasencia J, Churango Barreto K, Calvanapon Prado P, Salazar Abad S (2017) Prognostic significance of the ratio of lymph node metastatic in 5-year survival after curative gastrectomy for advanced gastric carcinoma. Rev Gastroenterol Peru **37**(3): 217–224.

40. Lu J, Wang W, Zheng CH et al (2017) Influence of Total Lymph Node Count on Staging and Survival After Gastrectomy for Gastric Cancer: An Analysis From a Two-Institution Database in China. Ann Surg Oncol **24**(2): 486–493.

41. Polkowski M, Palucki J, Wronska E, Szawlowski A, Nasierowska-Guttmejer A, Butruk E (2004) Endosonography versus helical computed tomography for locoregional staging of gastric cancer. Endoscopy **36**(7): 617–623.

42. Botet JF, Lightdale CJ, Zauber AG et al (1991) Preoperative staging of gastric cancer: comparison of endoscopic US and dynamic CT. Radiology **181**(2): 426–432.

43. Pereira MA, Ramos M, Dias AR et al (2018) Risk Factors for Lymph Node Metastasis in Western Early Gastric Cancer After Optimal Surgical Treatment. J Gastrointest Surg **22**(1): 23–31.

44. Chen L, Wang YH, Cheng YQ et al (2017) Risk factors of lymph node metastasis in 1620 early gastric carcinoma radical resections in Jiangsu Province in China: A multicenter clinicopathological study. J Dig Dis **18**(10): 556–565.

45. Miles KA (1999) Tumour angiogenesis and its relation to contrast enhancement on computed tomography: a review. Eur J Radiol **30**(3): 198–205.

46. Fukuya T, Honda H, Hayashi T et al (1995) Lymph-node metastases: efficacy for detection with helical CT in patients with gastric cancer. Radiology **197**(3): 705–711.

Affiliations

Chang Liu[1] · Liang Qi[2] · Qiu-Xia Feng[2] · Shu-Wen Sun[2] · Yu-Dong Zhang[2] · Xi-Sheng Liu[2]

[1] Department of Radiology, TongDe Hospital of ZheJiang Province, No. 234, Gucui Road, Hangzhou 310012, Zhejiang, China

[2] Department of Radiology, The First Affiliated Hospital of Nanjing Medical University, No. 300, Guangzhou Road, Nanjing 210009, Jiangsu, China