



OPEN Muscle-Driven prognostication in gastric cancer: A multicenter deep learning framework integrating iliopsoas and erector spinae radiomics for 5-Year survival prediction

Yuan Hong^{1,4}, Peng Zhang^{1,4}, Zhijun Teng^{1,4}, Kang Cheng², Zimo Zhang², Yixian Cheng¹, Guodong Cao¹✉ & Bo Chen^{1,3}✉

This study developed a 5-year survival prediction model for gastric cancer patients by combining radiomics and deep learning, focusing on CT-based 2D and 3D features of the iliopsoas and erector spinae muscles. Retrospective data from 705 patients across two centers were analyzed, with clinical variables assessed via Cox regression and radiomic features extracted using deep learning. The 2D model outperformed the 3D approach, leading to feature fusion across five dimensions, optimized via logistic regression. Results showed no significant association between clinical baseline characteristics and survival, but the 2D model demonstrated strong prognostic performance (AUC ~ 0.8), with attention heatmaps emphasizing spinal muscle regions. The 3D model underperformed due to irrelevant data. The final integrated model achieved stable predictive accuracy, confirming the link between muscle mass and survival. This approach advances precision medicine by enabling personalized prognosis and exploring 3D imaging feasibility, offering insights for gastric cancer research.

Keywords Sarcopenia, Gastric cancer, Radical gastrectomy, Deep learning

Abbreviations

Adam	Adaptive Moment Estimation
AI	Artificial Intelligence
AUC	Area Under the Curve
BIA	Bioelectrical Impedance Analysis
CAM	Class Activation Mapping
CNN	Convolutional Neural Network
CT	Computed Tomography
CUDA	Compute Unified Device Architecture
DCA	Decision Curve Analysis
DEXA	Dual Energy X-ray Absorptiometry
DenseNet	Densely Connected Convolutional Network
Grad-CAM	Gradient-weighted Class Activation Mapping
HIS	Hospital Information System
HIV	Human Immunodeficiency Virus
Inception	Inception Network

¹Department of General Surgery, The First Affiliated Hospital of Anhui Medical University, Hefei 230022, 218 Jixi Road, China. ²Department of The First Clinical Medical College, Anhui Medical University, Hefei 230022, China. ³The People's Hospital of Hanshan County, Ma'anshan 238101, China. ⁴These authors contributed equally to this work and should be considered co-first authors: Yuan Hong, Peng Zhang and Zhijun Teng. ✉email: ayfycgd@163.com; chenbo831116@163.com

IWGS	International Working Group on Sarcopenia
KNN	K-Nearest Neighbors
LightGBM	Light Gradient Boosting Machine
LR	Logistic Regression
MNASNet	Mobile Neural Architecture Search Network
MobileNet	Mobile Convolutional Neural Network
MRI	Magnetic Resonance Imaging
OpenCV	Open Source Computer Vision Library
PACS	Picture Archiving and Communication Systems
RECIST	Response Evaluation Criteria In Solid Tumors
ResNet	Residual Network
RF	Random Forest
ROC	Receiver Operating Characteristic Curve
ROI	Region of Interest
ShuffleNetV2	Shuffle Network Version 2
SPSS	Statistical Package for the Social Sciences
SVM	Support Vector Machine
TNM	Tumor, Node, Metastasis
VGG	Visual Geometry Group
WL	Window Level
WW	Window Width
XGBoost	eXtreme Gradient Boosting
ZPS	Zubrod Performance Status
2D	Two-Dimensional
3D	Three-Dimensional

Sarcopenia is markedly more prevalent in patients with gastric cancer and other cachexia-related conditions, and it is closely linked to patient prognosis. Patients with gastric cancer often experience progressive loss of skeletal muscle mass and function due to factors such as increased protein catabolism, chronic inflammation, metabolic disturbances, and impaired oral intake, leading to a cachectic state. This condition results in reduced treatment tolerance, elevated risk of postoperative complications, and poorer survival outcomes^{1–3}. Studies have demonstrated that sarcopenia is significantly associated with prolonged hospital stays, decreased sensitivity to chemotherapy, and lower 5-year survival rates in gastric cancer patients, with symptoms being especially pronounced in advanced stages⁴. Nutritional intake disorders induced by surgery or chemotherapy further exacerbate muscle wasting, thereby creating a vicious cycle⁵. Multiple clinical studies have clearly established that sarcopenia in gastric cancer patients is strongly correlated with increased postoperative complication rates, extended hospitalization, and diminished long-term survival^{6,7}. Therefore, early identification and intervention of sarcopenia are crucial to improving clinical outcomes by enhancing patients' quality of life, increasing treatment tolerance, and ultimately improving survival rates.

First introduced in the 1980s by Irwin Rosenberg, sarcopenia was originally defined as an age-related loss of skeletal muscle mass, particularly prominent among the elderly population⁸. With advancing research, the definition has evolved into a multidimensional clinical syndrome that encompasses not only reduced muscle mass but also declines in muscle strength and functional capacity. These three components interact synergistically to cause impaired mobility and diminished quality of life⁹. In 2022, the International Working Group on Sarcopenia (IWGS) further refined the definition, emphasizing sarcopenia as a comprehensive decline in muscle mass, strength, and physiological function associated with aging¹⁰. Importantly, sarcopenia is not merely an inevitable consequence of aging but also a critical independent factor impacting the health of older adults. It is closely linked to chronic conditions such as metabolic syndrome, cardiovascular disease, diabetes, and cancer, significantly elevating risks of mortality and complications^{11,12}. Current diagnostic criteria focus on multidimensional comprehensive evaluations, integrating imaging-based muscle mass assessments such as DEXA and CT scans, grip strength tests (reflecting muscle strength), and gait speed or 5-meter walking tests (assessing physical function)^{13,14}. This standardized system has been widely applied in global clinical and epidemiological studies and is regarded as the gold standard for sarcopenia diagnosis and management. European and Asian research organizations are actively promoting the unification of diagnostic criteria. By integrating multiple indicators, the efficiency of early identification and intervention for sarcopenia has significantly improved, providing a robust scientific basis for better patient outcomes. In addition to traditional assessments such as handgrip strength, gait speed, and imaging-based measurements, recent studies have developed interpretable logistic regression models and nonlinear XGBoost models for clinical stratification of sarcopenia risk. These approaches facilitate early identification of high-risk individuals, enabling personalized nutritional and rehabilitative interventions, thereby mitigating functional decline and improving long-term outcomes¹⁵. Moreover, studies incorporating the EWGSOP2 algorithm have highlighted bioelectrical impedance analysis (BIA) as a rapid, cost-effective, and clinically scalable method for evaluating muscle mass. BIA measurements show strong agreement with CT/DEXA imaging results, supporting its use as a bedside screening tool for large-scale population-based applications¹⁶. These emerging methodologies broaden the scope of sarcopenia research and intervention strategies, providing a practical foundation for the optimization of predictive models.

Radiomics, as an emerging quantitative medical imaging analysis technology, extracts high-dimensional features such as muscle volume, density, texture, and spatial distribution from CT, MRI, and other imaging modalities to analyze disease pathological mechanisms and biological characteristics^{17–19}. Compared with traditional imaging assessments, its advantage lies in providing objective and multidimensional biomarkers to

support disease diagnosis, prognosis prediction, and treatment decision optimization. In the field of sarcopenia, radiomics can precisely quantify muscle mass loss, metabolic abnormalities, and functional decline, enabling early screening, severity grading, and progression risk assessment²⁰. Furthermore, radiomics has demonstrated significant value in tumor prognosis studies, such as gastric cancer. By extracting features from patient imaging data related to muscle mass reduction, tumor aggressiveness, and immune microenvironment, predictive models can be constructed to evaluate postoperative complications, chemo-radiotherapy sensitivity, and survival rates. Studies have confirmed that radiomic biomarkers are closely associated with tumor heterogeneity, immune infiltration, and other biological behaviors, providing a basis for personalized treatment. By integrating imaging features of sarcopenia and tumors, treatment strategies can be optimized²¹. The application of radiomics is driving the precise development of sarcopenia and tumor research, offering non-invasive and efficient decision support for clinical practice²². With the continuous advancement of translational medicine, integrating molecular targets with radiomic features has emerged as a research hotspot. Studies have demonstrated that the neurotensin receptor, as a novel therapeutic target, holds significant clinical translational potential in gastric cancer. When combined with nanodrug delivery systems, it enables simultaneous therapy and image-guided interventions, thereby enhancing therapeutic efficacy while reducing systemic toxicity²³. In addition, radiomics approaches based on contrast-enhanced CT have recently been applied to preoperatively predict the differentiation status of gastrointestinal tumors²⁴. The extracted texture and morphological features from these models have shown promising performance in classifying tumor differentiation. Integrating such predictive techniques with traditional risk factors may provide a more comprehensive reference for preoperative assessment and treatment decision-making. These cutting-edge advances offer deeper support for the clinical application of radiomics in gastric cancer.

In recent years, deep learning technology has revolutionized medical imaging analysis, demonstrating exceptional performance in tasks such as image classification, segmentation, and feature extraction. Convolutional neural networks (CNNs) process single-layer image slices through 2D architectures, efficiently extracting local features, while 3D CNNs integrate spatial structural information from three-dimensional images, showcasing unique advantages in tumor morphology analysis, boundary identification, and heterogeneity assessment²⁵. The combination of deep learning and radiomics not only improves the accuracy of quantitative features such as muscle mass and texture but also facilitates the efficient analysis of pathological associations between sarcopenia and cancers. In the future, AI models integrating radiomics with multi-modal data such as genomics and metabolomics are expected to overcome the limitations of single data sources, driving the optimization of personalized treatment strategies.

The purpose of this study is to apply deep learning techniques, using postoperative 5-year survival status as the primary outcome, to conduct transfer learning with various deep neural networks and investigate the relationship between sarcopenia and prognosis in gastric cancer patients. This research aims to analyze muscle imaging data from gastric cancer patients to develop valuable predictive tools for clinical management and establish an efficient model for predicting 5-year survival rates. The findings are expected to enhance treatment outcomes and quality of life for gastric cancer patients, thereby advancing the field of personalized medicine.

Methods and materials

The establishment of the clinical cohort

This retrospective radiomics study, approved by the Ethics Committee of the First Affiliated Hospital of Anhui Medical University (PJ2025-02-33), was conducted in accordance with the ethical principles of the Declaration of Helsinki. All participants provided written informed consent for the use of their clinical data and preoperative CT images in research. To ensure privacy protection, all personally identifiable information was removed from the imaging and clinical datasets prior to analysis. The study involved a retrospective analysis of clinical data from 824 patients diagnosed with gastric cancer. The diagnosis dates ranged from December 2018 to June 2019. Strict inclusion and exclusion criteria were established to ensure the accuracy and reliability of the research results. The criteria were as follows:

- a. Patients aged 18 years or older, with no restrictions on gender;
- b. Patients with clearly evaluable lesions (RECIST 1.1) who were clinically assessed as candidates for elective gastric resection and histopathologically diagnosed with gastric adenocarcinoma;
- c. Patients with no prior history of immunotherapy or chemotherapy for metastatic lesions, but those who underwent radical surgery and postoperative adjuvant therapy were eligible;
- d. Zubrod-ECOG-WHO (ZPS, 5-point scale) performance status score of 1;
- e. No active autoimmune diseases, active hepatitis B, or HIV infection;
- f. Patients with good bone marrow, liver, kidney, heart, lung, and nervous system function as assessed by clinicians, and who were deemed capable of tolerating chemotherapy, immunotherapy, and surgical treatment;
- g. Patients willing to participate in and cooperate with all aspects of the study, including but not limited to treatment, follow-up, data collection by researchers, and refraining from other treatments.

The patient enrollment process is shown in Fig. 1. A total of 603 patients who underwent radical gastrectomy were screened and included in the follow-up. During the follow-up, 32 patients voluntarily withdrew, 57 were lost to follow-up, and 3 deaths unrelated to the study were recorded. Ultimately, five-year survival data for 511 patients were collected and analyzed. To validate the model and reduce overfitting, preoperative CT and clinical data from 194 gastric cancer patients treated at Hanshan County People's Hospital between 2017 and 2019 were used as an external validation cohort. The inclusion criteria were consistent, and all patients completed five-year follow-up. Key survival data were collected to evaluate the model's performance.

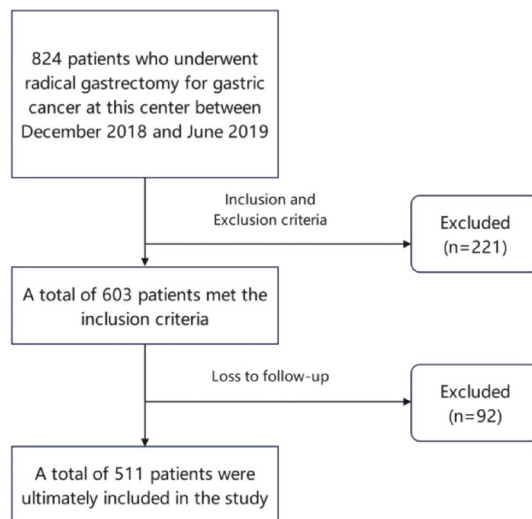


Fig. 1. Flowchart of patient enrollment. Of 603 initially screened gastrectomy patients, 511 completed 5-year follow-up (32 withdrew, 57 lost to follow-up, 3 unrelated deaths).

In this study, CT data were obtained from routine preoperative scans performed using a 256-slice scanner. The image parameters included a 512×512 matrix with a resolution of 0.62–0.86 mm. The window width (WW) and window level (WL) were fixed at 2000 and –1000, respectively. This setting minimizes unnecessary noise and artifacts while maximizing the visualization of soft tissue structures and pathological changes. To eliminate variability caused by equipment differences or operational deviations, the images were standardized through preprocessing steps such as alignment, denoising, and enhancement. These processes improved image quality and provided a reliable basis for diagnostic analysis.

Clinical data statistical analysis

This study involved 11 basic clinical indicators, including variables such as height, weight, and TNM staging. Detailed statistical analyses were conducted on these indicators during the collection of clinical baseline data. The research team performed a comparative analysis of baseline clinical data to verify the consistency between the data from this center and the external dataset. In the survival analysis phase, univariate Cox regression analysis was employed to identify potential factors influencing survival. Variables with a *p*-value less than 0.05 in the univariate analysis were included in the multivariate Cox regression model to further identify factors with independent prognostic value for survival. In the multivariate Cox regression model, clinical factors confirmed as independent prognostic factors showed statistically significant differences ($p < 0.05$). Subsequently, a clinical dataset containing multimodal integrated information was constructed, with strictly selected independent clinical prognostic factors extracted for further analysis.

Muscle modeling and segmentation

The muscle delineation process is shown in Fig. 2. The original CT images were used as the baseline, and the TotalSegmentation software, based on the U-NET deep learning model, was employed for automated segmentation. The model was implemented on the PyTorch 1.11 deep learning framework, running with CUDA version 11.3. Training and inference were conducted on an NVIDIA GeForce RTX 4090 GPU. The training parameters were configured as follows: learning rate of 0.001, batch size of 32, and 100 iterations.

All generated muscle models were independently reviewed by two experienced radiologists. They meticulously examined all delineation results and manually corrected any deviations from standard anatomical structures to ensure the accuracy of the automated segmentation. In addition, the original abdominal CT images were reprocessed, and regions of interest (ROI) models were automatically delineated to ensure consistency in spatial resolution across all datasets. This enabled images from different sources to be compared on the same scale. The voxel size of the images was standardized to $3 \times 3 \times 3 \text{ mm}^3$. The AutomatedCut tool (Version 2.0, based on the OpenCV library) was used to segment the abdominal CT images and their corresponding ROI regions. The maximum cross-sectional slice of each muscle was selected, as it typically reflects the largest muscle area and consistent anatomical landmarks. This method is widely adopted in muscle quantification studies due to its strong correlation with total muscle volume and high reproducibility across individuals²⁶. The L3 vertebral level, in particular, serves as a validated reference point for such assessments in abdominal CT scans²⁷. Accordingly, maximum cross-sectional images were extracted to ensure consistency and comparability across the dataset. The maximum cross-sectional images for each muscle modality were extracted, and the muscle models were divided into three groups: bilateral iliopsoas muscles, bilateral erector spinae muscles, and a combined modality group. The research team used a window width and window level adjustment tool to modify the CT images. From both the original CT images and the adjusted abdominal window images, the maximum cross-sectional images of the three muscle modalities were obtained. This process successfully generated six distinct input datasets, covering

Automatic Segmentation Flow Chart

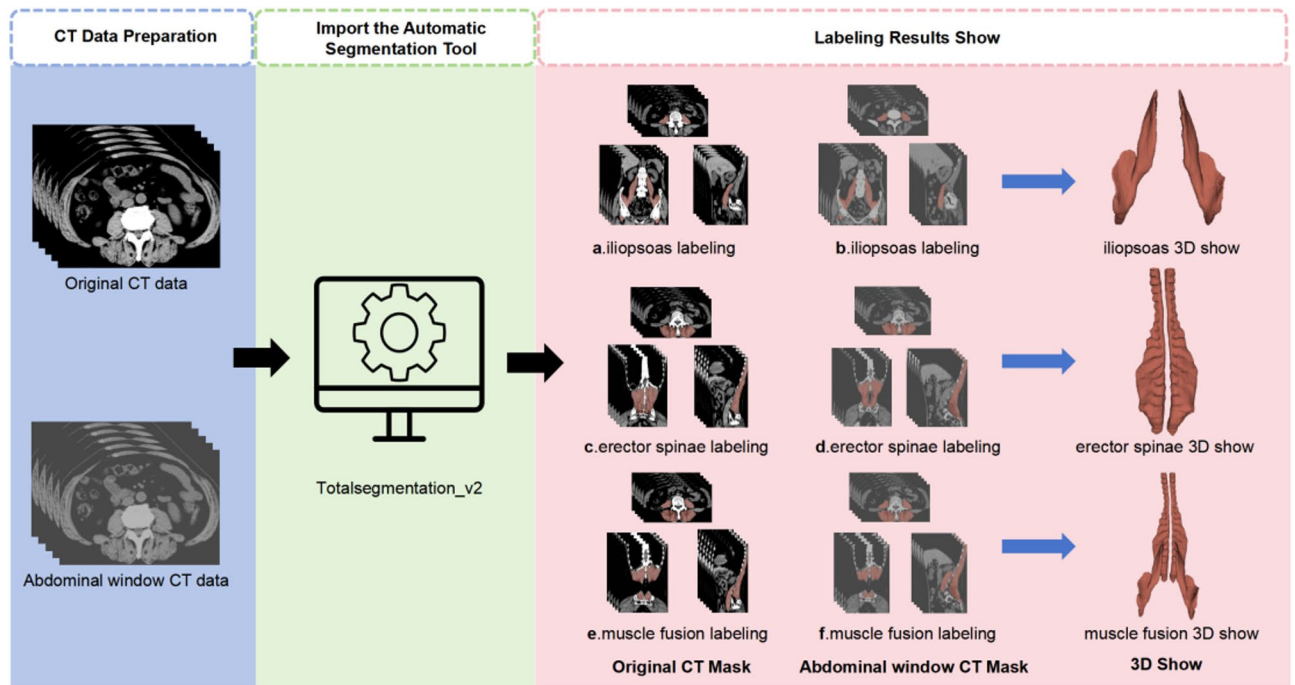


Fig. 2. Flow chart of muscle segmentation. The pipeline processes abdominal CT images through three stages: (1) original/abdominal window phases, (2) automated segmentation using a U-NET model (TotalSegmentation), and (3) generation of three muscle subgroups (bilateral psoas, erector spinae, and combined muscles). Six standardized input datasets (3 muscle groups $\times$ 2 CT phases) were produced for subsequent analysis.

the maximum cross-sectional images of the bilateral iliopsoas muscles, bilateral erector spinae muscles, and their combined modality under both the original and abdominal windows.

Deep radiomics workflow

In this study, radiomic features refer solely to deep learning–derived features extracted from the penultimate layer of convolutional neural networks. Handcrafted features from traditional methods (e.g., PyRadiomics) were not included. All references to radiomics in this paper denote deep features, unless otherwise noted.

2D neural network training workflow

The radiomics workflow of this research is shown in Fig. 3. After obtaining six different types of input data, the research team trained models using various neural network architectures. From the ResNet family, ResNet18, ResNet50, and ResNet152 were selected; from the DenseNet family, DenseNet121 and DenseNet201 were chosen; from the Inception family, GoogLeNet was used; from the MNASNet family, MNASNet0.5 and MNASNet1.0 were selected; from the MobileNet family, MobileNet_v3_large and MobileNet_v3_small were included; from the ShuffleNetV2 family, ShuffleNetV2_x0.5 and ShuffleNetV2_x1.0 were adopted; and from the VGG family, VGG13_bn and VGG19_bn were utilized. A total of 14 deep learning models were implemented to leverage the unique strengths of each network. This approach ensured an optimal balance of performance, computational efficiency, and applicability in building the predictive models, providing robust technical support for precision medicine and personalized treatment.

We selected data from 511 patients as the training cohort and data from 194 patients as an external validation cohort. Multiple deep learning models were trained independently, each for 100 epochs. During training, the Adam optimizer was uniformly applied with an initial learning rate of 0.001, which decayed to one-tenth of its original value every 30 epochs. This learning rate schedule was designed to accelerate convergence and reduce the risk of overfitting. A batch size of 32 was used, and binary cross-entropy was adopted as the loss function. The training cohort was randomly split into training and internal validation subsets at a 1:4 ratio.

The above hyperparameter configuration was empirically fine-tuned based on multiple rounds of experimental results, balancing computational cost with model stability. This setup yielded satisfactory convergence and generalization performance across all network architectures. Given that this study employed classical neural network structures pre-trained on ImageNet (e.g., ResNet, DenseNet), and previous studies have shown that modest fine-tuning can yield excellent performance in transfer learning scenarios^{28,29}, we retained the original network architectures (e.g., number of layers, convolutional kernel parameters) and only adjusted optimization-related parameters as needed.

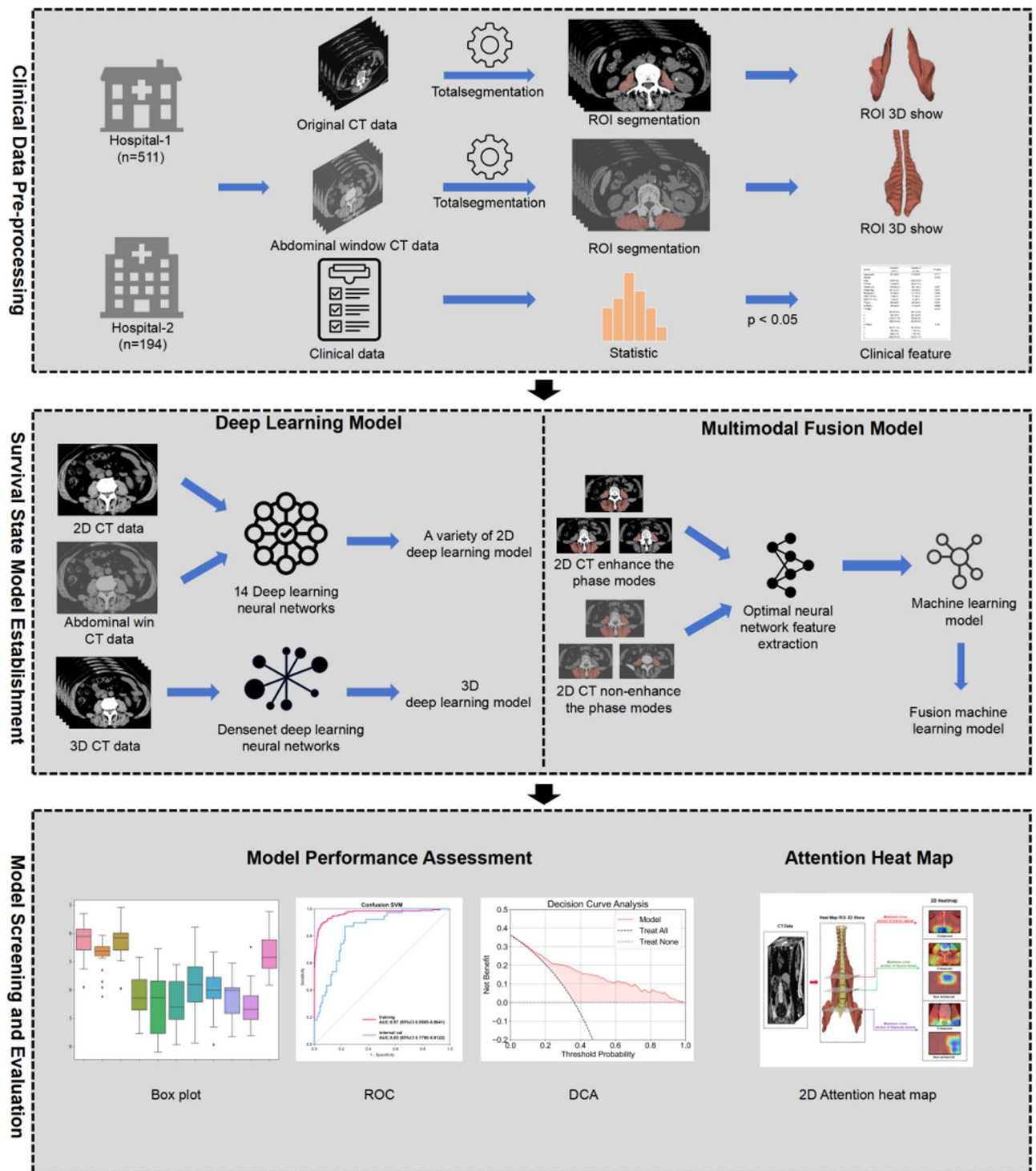


Fig. 3. Radiomics workflow for 5-year survival prediction after gastrectomy. The pipeline consists of three main stages: (1) preprocessing of clinical and imaging data, (2) multi-modal 2D neural network screening (14 architectures tested) with feature fusion, and (3) predictive model evaluation using 11 machine learning classifiers. The optimal 2D-based fusion strategy demonstrated superior performance over 3D approaches in external validation ($n = 194$).

After model training, ROC curves were plotted, and the AUC was calculated to quantify classification performance. In addition, accuracy, sensitivity, specificity, and precision were also evaluated to comprehensively assess each model's performance from multiple perspectives and determine its potential applicability in clinical practice.

3D neural network training workflow

After a comprehensive evaluation of the classification performance of traditional 2D neural networks, the research team decided to employ the application of 3D neural networks. Compared to 2D neural networks, 3D neural networks can leverage richer spatial information and handle more complex volumetric image data, offering the potential to enhance the accuracy and robustness of medical image analysis. Unlike 2D neural networks, which only process single-plane data, 3D neural networks can receive and process complete three-dimensional image data. This allows them to fully capture the spatial characteristics of organs and lesions in the depth dimension, thereby improving their ability to identify complex disease structures. For these reasons, the research team concluded that 3D neural networks are better suited for handling complex clinical medical imaging data, particularly in tasks such as tumor detection and organ segmentation, where they may demonstrate significant advantages. To this end, the research team performed rigorous preprocessing and normalization of the patients' complete CT data, ensuring consistency across datasets. All input image phases were standardized to the abdominal contrast-enhanced phase. Subsequently, DenseNet121 and DenseNet201 were selected as the two 3D neural network models for training. The training environment and parameter settings were kept consistent with those used in the earlier 2D neural network studies.

Multimodal data fusion diagnostic model based on 2D neural networks

Through systematic architectural comparison experiments, the research team observed that 3D neural networks did not demonstrate significant advantages under current medical imaging data characteristics. After comprehensive evaluation of empirical evidence and computational resource efficiency, 2D neural networks were ultimately adopted to achieve more robust predictive capabilities. The best-performing 2D neural network for each modality was selected, as shown in Table 1. Ultimately, five optimal 2D deep learning models from different modalities were chosen for multimodal fusion. A feature fusion strategy was employed by extracting intermediate representations from the penultimate layer (namely the pooling layer) of each selected neural network classifier. The dimensionality of the extracted features varied according to the architecture: 512 for ResNet18, 128 for MNASNet0_5, and 2048 for both ResNet152 and ResNet50. The multimodal fused feature vector was constructed by concatenating the outputs from the five selected models, resulting in a final feature vector with a total dimensionality of 5248. This high-dimensional representation was subsequently used as the input for downstream machine learning classifiers.

After completing the data transformation, the research team employed 11 machine learning algorithms as final classifiers for the fused features. These included methods such as Support Vector Machines (SVM), Random Forest (RF), K-Nearest Neighbors (KNN), XGBoost, and LightGBM. To further optimize the above classifiers, we conducted a systematic hyperparameter tuning process using a grid search strategy. Key parameters—including learning rate, number of trees, maximum depth, and regularization coefficients—were adjusted based on performance on the training and internal validation sets to identify the optimal configuration. All models were trained and evaluated under a consistent data split and cross-validation framework to ensure the comparability and robustness of performance across classifiers. After a comprehensive evaluation considering both predictive accuracy and model interpretability, logistic regression was ultimately selected as the output classifier for the multimodal fusion model. The regularization parameter (C-value) of the logistic regression model was further fine-tuned to enhance classification performance. This multimodal fusion strategy yielded a high-precision predictive model for assessing five-year postoperative survival in gastric cancer patients, demonstrating promising potential for clinical application.

Attention heatmaps visualization

Attention heatmaps, as a visualization tool in deep learning, are widely applied in computer vision tasks to intuitively showcase key regions involved in neural network decision-making, thereby enhancing model interpretability. In medical imaging analysis, attention heatmaps provide supplementary information for clinical experts, improving the model's reliability and usability. This study employs 2D attention heatmap techniques combined with the Grad-CAM module to deeply analyze the areas of focus in neural networks when predicting the survival rates of gastric cancer patients. Grad-CAM generates class activation maps based on gradient calculations, making it applicable to various deep learning architectures without requiring modifications to the network structure.

The research team input the trained 2D deep learning model with test images and used Grad-CAM to compute activation maps, generating heatmaps of the same dimensions as the input images. Regions with high intensity highlight areas that contribute most to the decision-making process. To enhance visualization, the heatmaps were overlaid on the original medical images and enhanced with pseudo-color mapping. This method allows for

Modality	Model
2D Non-enhanced muscle fusion	resnet18
2D Enhanced muscle fusion	resnet18
2D Non-enhanced iliopsoas	mnasnet0_5
2D Enhanced iliopsoas	resnet152
2D Non-enhanced erector spinae	—
2D Enhanced erector spinae	resnet50

Table 1. Mode fusion statistics table.

a systematic analysis of the model’s focus, evaluates its rationality and robustness, and identifies potential biases, such as over-reliance on irrelevant features or neglect of key anatomical regions.

2.6.Statistic analysis

The research team used SPSS 27.0 software for data processing and analysis. Categorical variables were analyzed using the chi-square test, while continuous variables were tested with either an independent samples t-test or the Mann-Whitney U test, depending on the distribution characteristics of the data. Survival analysis was performed using univariate and multivariate Cox proportional hazards models.

Radiomics analysis was conducted in Jupyter Notebook using the Python programming language. Libraries such as scikit-learn and pandas were utilized for feature selection, data preprocessing, and dimensionality reduction. To facilitate the presentation of results and data visualization, the research team used the matplotlib library to generate high-quality imaging and analytical charts. This study was conducted using the Onekey platform for model training. The source code is publicly available at: <https://gitee.com/wangqingbaidu/OnekeyCompo>.

3.Results

Clinical baseline and survival analysis statistical results

Thirteen clinical variables were collected and compared between the internal and external cohorts (Table 2). The results showed that only serum albumin levels differed significantly between groups. The distribution of continuous variables was assessed using the Shapiro-Wilk test. Binary variables were analyzed with Fisher’s exact test or Pearson’s chi-square test (with continuity correction), while multicategorical variables were assessed via contingency table chi-square tests. Among continuous variables, height was analyzed by independent samples t-test, whereas age, weight, and BMI were analyzed using the Mann-Whitney U test. All tests were two-tailed, with significance set at $P < 0.05$.

In survival analysis, univariate Cox regression identified weight, BMI, and red blood cell count as significant factors affecting survival ($P < 0.05$). However, multivariate Cox regression revealed that none of these factors independently predicted 5-year survival in gastric cancer patients (Table 3).

Performance of 2D deep learning models

Transfer learning experiments were conducted on 14 neural network models across six modalities, with their predictive performances systematically compared. Models with accuracy and AUC below 0.5 on both training and validation sets were excluded due to poor classification performance. Among the remaining models, those demonstrating superior classification accuracy and low overfitting risk were selected as optimal for each modality.

Figure 4 presents ROC and DCA curves for the best-performing 2D neural networks per modality, with detailed training results provided in Appendix Table 1. For the combined muscle phases, ResNet18 achieved consistent accuracy and AUC values above 0.6 in both training and test sets, indicating robustness and minimal overfitting; thus, it was selected as the optimal model for both the original and abdominal enhancement phases.

Variants	Hospital-1 (n = 511)	Hospital-2 (n = 194)	P-value
Age (years)	62.3 ± 9.8	61.9 ± 9.9	0.717
Sex			0.794
Male	378(74%)	146(75.3%)	
Female	133(26%)	48(24.7%)	
Height (cm)	165.8 ± 6.4	166.1 ± 6.2	0.601
Weight (kg)	60.1 ± 10.1	59.9 ± 9.5	0.857
BMI (kg/m²)	21.8 ± 3.3	21.7 ± 3.3	0.846
WBC (10^9/L)	5.9 ± 2.3	5.7 ± 2.4	0.473
RBC (10^12/L)	4.2 ± 0.6	4.2 ± 0.7	0.206
Total Protein (g/L)	66.6 ± 6.5	66.5 ± 6.5	0.927
Albumin (g/L)	40.5 ± 4.6	41.2 ± 4.6	0.046
T-stage			0.232
1	95(18.6%)	28(14.4%)	
2	40(7.8%)	25(12.9%)	
3	210(41.1%)	78(40.2%)	
4	166(32.5%)	63(32.5%)	
N-stage			0.39
0	62(12.1%)	32(16.5%)	
1	18(3.5%)	11(5.7%)	
2	26(5.1%)	11(5.7%)	
3	405(79.3%)	140(72.1%)	

Table 2. Clinical baseline statistics table.

Variants	Univariate Cox		Multivariate Cox
	HR (95%CI)	P-value	P-value
Age (years)	1.00(0.99,1.02)	0.807	
Sex		0.953	
Male	Ref		
Female	1.01(0.69,1.49)	0.953	
Height (cm)	0.99(0.96,1.01)	0.287	
Weight (kg)	0.97(0.95,0.98)	<0.001	0.328
BMI (kg/m ²)	0.91(0.86,0.96)	0.001	0.757
WBC (10 ⁹ /L)	0.96(0.89,1.04)	0.305	
RBC (10 ¹² /L)	0.72(0.56,0.93)	0.01	0.129
Total Protein (g/L)	0.99(0.96,1.01)	0.374	
Albumin (g/L)	0.99(0.94,1.01)	0.218	
T-stage		0.403	
1	Ref		
2	0.81(0.47,1.40)	0.456	
3	0.62(0.31,1.22)	0.168	
4	1.04(0.71,1.53)	0.831	
N-stage		0.184	
0	Ref		
1	0.60(0.34,1.05)	0.073	
2	0.68(0.28,1.68)	0.403	
3	0.57(0.23,1.40)	0.217	

Table 3. Cox survival analysis statistics table.

In the iliopsoas muscle phase, DenseNet121 and MNASNet0.5 showed comparable performance with low overfitting risk in the original phase, but MNASNet0.5 was chosen to enhance model heterogeneity. Its significant classification performance was confirmed by the DCA curve. ResNet152 outperformed other models in the abdominal enhancement phase with stable results across datasets, earning its selection as the optimal model.

For the erector spinae muscle phase, models performed poorly in the original phase with accuracies near random chance (0.5), leading to abandonment of training for this phase. However, ResNet50 demonstrated the best performance in the abdominal enhancement phase despite some overfitting risk, supported by a significant AUC confidence interval.

Performance of 3D deep learning models

To investigate the performance of 3D neural networks in medical image classification, this study selected two advanced 3D architectures from the DenseNet family, DenseNet121 and DenseNet201, due to their excellent feature extraction capabilities enabled by their densely connected design, making them well-suited for complex three-dimensional medical imaging data. The experiments directly utilized complete preoperative CT scans of patients as input, with a specific focus on abdominal enhancement phase CT images to improve model recognition accuracy. However, the experimental results fell short of expectations, with both models achieving AUC values below 0.5, indicating significant challenges in the classification task (Fig. 5; Table 4). Although 3D images provide richer spatial information and structural detail, the poor model performance may be attributed to a mismatch between the architectures and the dataset or interference from data noise. As a result, this study temporarily excluded 3D neural networks from multimodal integration. Future research will focus on optimizing the models and preprocessing the data, exploring additional 3D architectures and applying data cleaning techniques to fully leverage the diagnostic potential of 3D imaging, thereby advancing medical image classification technology.

Performance of multimodal fusion models

During the model fusion process, we extracted the penultimate layers, specifically the pooling layer outputs, from all the selected 2D neural network models. As a critical component in neural networks, pooling layers effectively preserve important features while reducing data dimensionality, and their output data typically contain rich feature information. After successfully obtaining these pooling layer outputs, we performed the necessary preprocessing steps to transform the raw 2D image data into a format suitable for further analysis. Then, we employed concatenation techniques to effectively integrate the pooling layer outputs from different neural network models. This step aimed to fuse the feature information from multiple models, thereby enhancing the performance of subsequent classification tasks. We then selected 11 mainstream machine learning algorithms as classifiers for performance evaluation (refer to Table 2). Keeping the original data grouping unchanged, we conducted 50-fold cross-validation and identified that the Logistic Regression (LR) and Support Vector Machine (SVM) models performed the best on the fused features (Fig. 6). Considering both the generalization

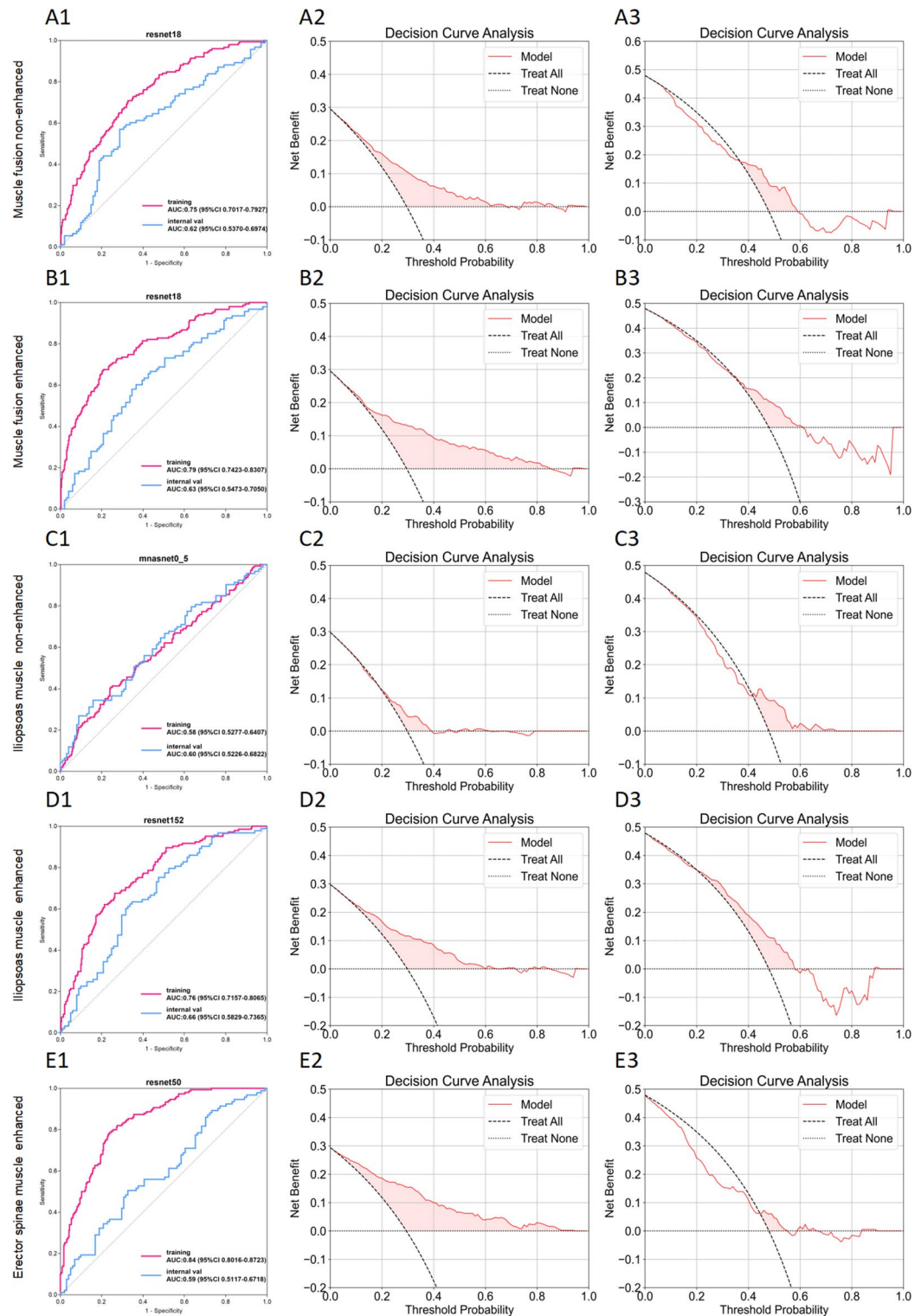


Fig. 4. Performance evaluation of optimal 2D neural networks across five imaging modalities. (A-E) The receiver operating characteristic (ROC) and decision curve analysis (DCA) results demonstrate the predictive performance of selected models: a. Muscle fusion: ResNet18 for both non-enhanced (A) and enhanced (B) modalities (AUC > 0.6). b. Iliopsoas: MNASNet0_5 (C, non-enhanced) and ResNet152 (D, enhanced). c. Erector spinae: ResNet50 (E, enhanced only; original phase models excluded due to accuracy $\approx$ 0.5). d. Each modality is presented in three subplots: (1) ROC curves with AUC and 95% CI, (2) training set DCA, and (3) test set DCA. Model selection was based on strict criteria excluding underperforming models (AUC < 0.5) and prioritizing those with consistent train-test performance and low overfitting risk.

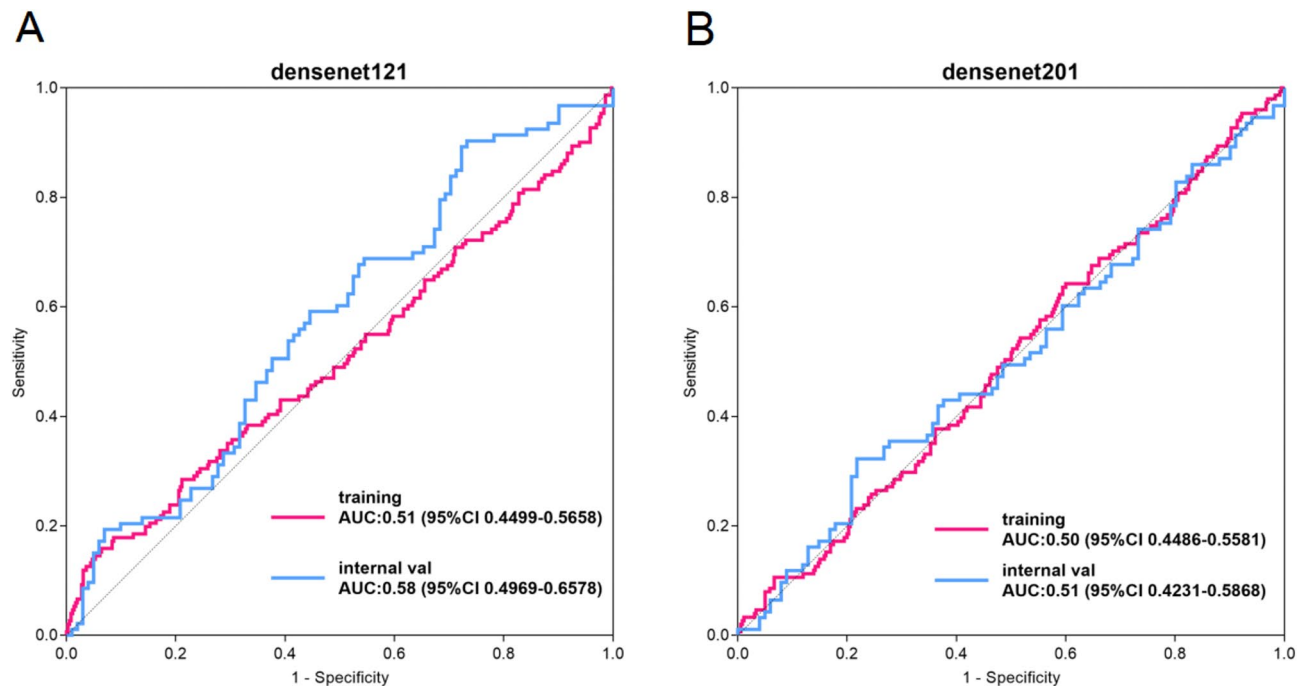


Fig. 5. Performance comparison of 3D neural networks using ROC curves. The evaluation of DenseNet121 and DenseNet201 models on abdominal CT scans showed limited classification capability, with AUC values below 0.5 for both models. While 3D architectures can capture spatial information effectively, these results suggest they may not be optimal for this particular task. The ROC curves (left: DenseNet121; right: DenseNet201) with 95% confidence intervals demonstrate this performance limitation, leading to their exclusion from the final model integration.

Model names	Accuracy	AUC	95% CI	Task
densenet121	0.701	0.492	0.4340–0.5501	Train
densenet121	0.531	0.423	0.3425–0.5038	Test
densenet201	0.481	0.497	0.4420–0.5517	Train
densenet201	0.5	0.495	0.4129–0.5770	Test

Table 4. Training results of the 3 D neural network.

performance of the models and the stability requirements for practical applications, we ultimately selected the LR model as our final fusion model.

3D attention heatmaps based on neural networks

To gain a deeper understanding of the model's decision-making process and its underlying biological significance, we utilized attention heatmap techniques to further analyze the best-performing neural network in each 2D modality (see Fig. 7). During classification, the neural network primarily focused its attention on the muscle tissue surrounding the spinal bones in the medical images. In addition to analyzing correctly predicted cases, we also performed attention heatmap visualization for misclassified samples. As shown in Fig. 8, correctly predicted cases exhibited concentrated attention regions primarily located in clinically relevant muscle structures, such as the iliopsoas and erector spinae muscles. In contrast, misclassified cases showed dispersed or deviated attention patterns, with the model focusing more on background or non-muscular areas.

This finding not only highlights the efficiency and accuracy of neural networks in processing medical imaging data but also provides a novel perspective for examining the relationship between muscle mass and patient prognosis. Based on this significant discovery, we plan to conduct further research on the correlation between muscle mass and prognosis in gastric cancer patients. By analyzing patients' muscle mass data in detail and integrating the predictive results of neural networks, we aim to provide more precise prognostic assessments and personalized treatment plans for patients.

Discussion

This study investigates the critical role of sarcopenia in predicting the prognosis of gastric cancer patients by integrating deep learning techniques with radiomics analysis. The proposed five-year postoperative survival

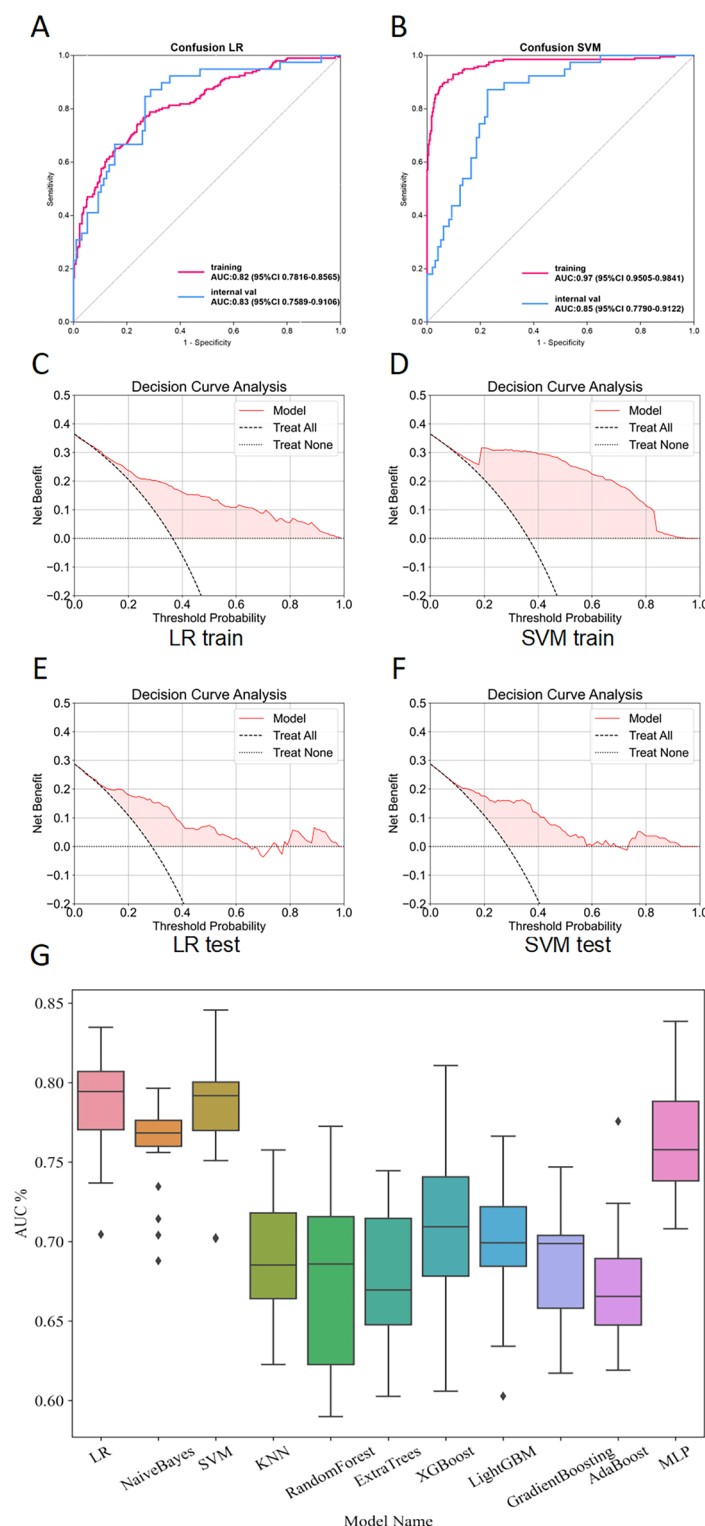


Fig. 6. Performance comparison of machine learning classifiers after multimodal feature fusion. The evaluation of 11 classifiers on fused neural network features showed logistic regression (LR) and support vector machine (SVM) as top performers. (A-B) ROC curves demonstrate classification accuracy of LR and SVM models. (C-F) DCA curves validate clinical utility in both training and test sets. (G) Boxplot summarizes cross-validated accuracy across all classifiers. The final selection of LR as optimal classifier was based on its balanced performance and practical stability in clinical application scenarios.

2D Heat Map Based on Neural Network

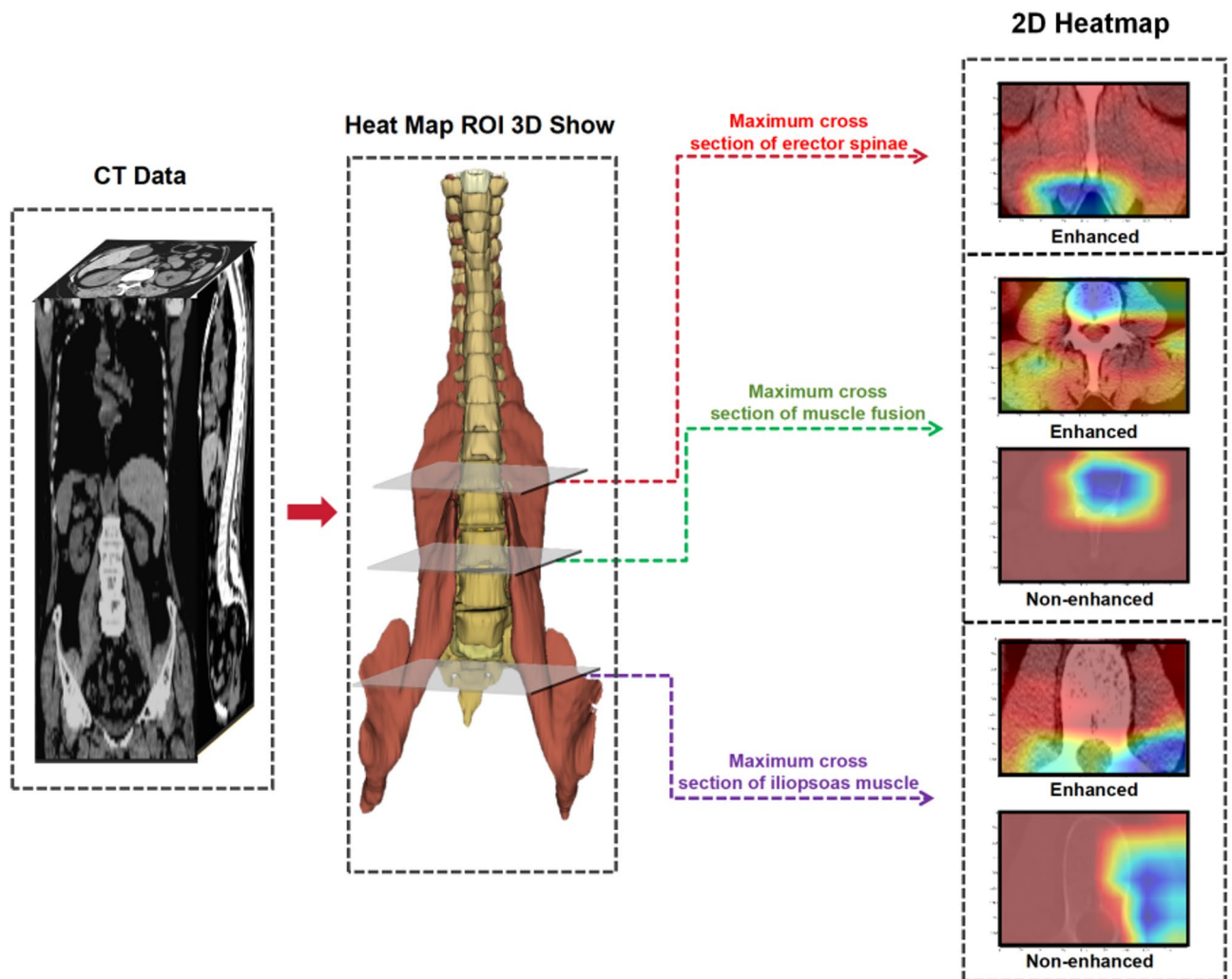


Fig. 7. 2D attention heatmaps based on neural networks. Models consistently focused on paravertebral muscles during classification, shown for (A) erector spinae, (B) iliopsoas, and (C) combined muscle groups.

prediction model offers new insights and technical support for individualized treatment strategies. Through a retrospective analysis of preoperative CT images and clinical data, the study validated the close association between sarcopenia and the prognosis of gastric cancer patients and demonstrated the potential of deep learning-based radiomics techniques in the prognostic evaluation of complex diseases. Furthermore, this study is the first to explore the use of multimodal neural network fusion in gastric cancer prognosis prediction models, optimizing model performance and paving a new path for the radiological diagnosis of sarcopenia and precision medicine for gastric cancer patients.

In recent years, radiomics has made significant progress in predicting the prognosis of gastric cancer³⁰. Previous studies have demonstrated that radiomic features derived from CT, MRI, and PET imaging can be used to evaluate the postoperative survival of gastric cancer patients³¹. For instance, some studies have successfully predicted lymph node metastasis in gastric cancer patients using texture analysis methods based on CT radiomics^{32,33}, which were further applied for prognostic stratification. Additionally, multiple studies have shown that radiomic features can be used to assess the efficacy of preoperative neoadjuvant therapy in gastric cancer and predict patients' long-term survival³⁴. However, most radiomics studies remain limited to traditional statistical methods, with the application of deep learning technology in this field still in its early stages³⁵. Building on previous work, this study introduces deep learning methods to enhance the performance of prediction models.

To comprehensively evaluate the predictive performance of our proposed model, we compared it against several representative studies published in recent years. Our multimodal fusion framework integrates five independently trained 2D deep learning subnetworks, each based on distinct muscle regions. The outputs of these subnetworks were aggregated using logistic regression. On the training cohort, the model achieved an

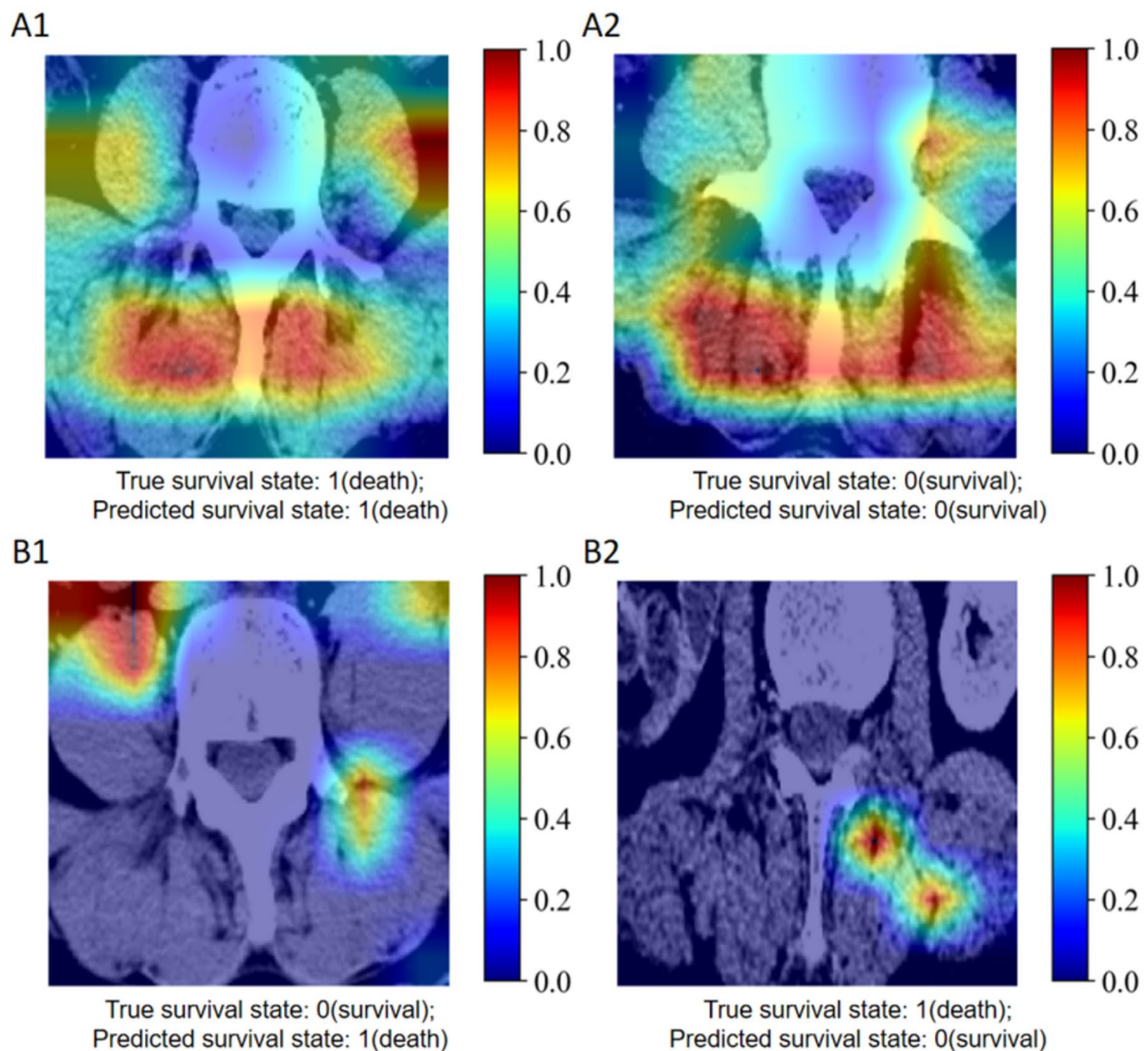


Fig. 8. Visualization of attention heatmaps for correctly and incorrectly predicted cases. (A) Heatmaps from correctly predicted cases demonstrate model focus on clinically relevant muscle areas. (B) Heatmaps from misclassified cases show scattered or non-informative attention regions.

AUC of 0.819 (95% CI: 0.7816–0.8565), and demonstrated robust generalization with an AUC of 0.835 (95% CI: 0.7589–0.9106) on an external test cohort from a different medical center.

In comparison, Chen et al. developed a skeletal muscle deep learning model based on a Transformer architecture to predict postoperative survival in gastric cancer patients. Their model yielded AUCs of 0.876 and 0.859 for recurrence-free survival and disease-specific survival on the internal validation cohort, and 0.823 and 0.822 on the external cohort, respectively—significantly outperforming traditional sarcopenia indicators (AUC range: 0.382–0.624, DeLong test $P < 0.001$)³⁶. Xiang et al. constructed a radiomics risk score model using contrast-enhanced CT features of the tumor region. Although the model achieved a training cohort AUC of 0.835 (95% CI: 0.793–0.877) for overall survival, its performance dropped to 0.730 (95% CI: 0.638–0.822) on the validation set; a similar decline was observed in disease-free survival prediction (training: 0.733; validation: 0.695), suggesting limitations in tumor-based features for long-term prognostic modeling³⁷. Hao et al. proposed the SurvivalCNN model, which combined 3D CT imaging with clinical variables, achieving C-indices of 0.849 for overall survival and 0.783 for progression-free survival in a cohort of 1,061 patients, outperforming multiple benchmark models and further supporting the effectiveness of multimodal feature fusion³⁸.

While our model demonstrated slightly lower performance on some individual metrics compared to Transformer-based or 3D models, it consistently achieved high accuracy across datasets from two independent centers. This consistency underscores its strong generalizability and adaptability across clinical settings. Overall, our 2D multimodal deep learning model based on muscle region features demonstrates competitive predictive

power, along with advantages in interpretability, computational efficiency, and stability when applied to moderately sized datasets—underscoring its potential for clinical translation.

In the clinical cohort analysis, we found that patients from the two centers exhibited significant differences only in albumin levels, which may be attributed to sampling bias and regional factors rather than being the primary focus of this study. In both univariate and multivariate Cox survival analyses, we found no significant association between the collected clinical data and the postoperative survival status of gastric cancer patients. This result is consistent with other clinical cohort studies, indicating that BMI, surgical methods, and laboratory indicators are insufficient for effectively predicting postoperative prognosis. This highlights the challenges in managing patients after gastric cancer surgery.

In contrast, the neural network trained on patients' muscle cross-sectional areas was able to directly predict the postoperative survival time of gastric cancer patients to a certain extent. After multimodal fusion, the model provided clinically valuable prediction results, further validating the potential of combining radiomics with deep learning. The survival analysis results also confirmed the clinical value of the model. However, this study only utilized four metabolomics features—white blood cell count, red blood cell count, albumin, and total protein—leaving room for improvement in data dimensionality. Previous studies have shown that metabolomic features play a significant role in cancer prognosis evaluation³⁹. Therefore, future research could further explore metabolomic data to identify key metabolic factors that significantly impact gastric cancer prognosis, thereby enhancing the performance of the prediction model.

This study demonstrates that 2D deep learning models exhibit stable and superior performance. Transfer learning experiments showed that ResNet18 achieved high accuracy and AUC values in both the iliopsoas muscle and abdominal enhanced phases, with similar performance between the training and testing sets, indicating a low risk of overfitting. Additionally, the MNASNet0_5 model performed well in the iliopsoas muscle phase and demonstrated high clinical utility in the decision curve analysis (DCA). In contrast, the neural network performance in the erector spinae muscle phase was suboptimal, particularly under the original phase conditions, where none of the 14 neural networks trained achieved satisfactory predictive models. This suggests that the contribution of different muscle groups to gastric cancer prognosis prediction varies. When selecting 2D cross-sectional planes, combining the iliopsoas and erector spinae planes may provide better predictive performance.

In this study, we integrated multiple neural networks with superior classification performance to enhance the predictive capability of our model. As a result, we achieved an AUC of approximately 0.8 on both the training and validation cohorts. Given that the input data were preoperative CT scans, this level of performance holds significant clinical value and may assist in formulating personalized treatment strategies.

However, the 3D neural networks did not outperform their 2D counterparts in predictive accuracy, a finding that drew considerable attention from our research team. Theoretically, 3D imaging captures more comprehensive spatial and structural information than 2D imaging and is thus expected to confer stronger predictive power. Contrary to expectations, our results demonstrated no significant performance gain with 3D models under the current configuration. Upon further investigation, several critical factors may account for this limitation.

First, during data preprocessing, although ROIs encompassing muscle tissue were selected, high-precision segmentation was not performed. As a result, adjacent non-muscle tissues were inadvertently included, introducing feature contamination and background noise. Second, despite employing two well-established 3D architectures, both exhibited suboptimal performance under standard parameter settings, with AUC values falling below 0.5. This suggests that the models failed to learn meaningful features, indicating that the performance bottleneck likely stems not only from network architecture or hyperparameter choices, but more fundamentally from imprecise ROI delineation and suboptimal adaptation of 3D feature extraction methods^{40,41}.

Moreover, existing 3D convolutional neural networks face inherent challenges in handling high-dimensional medical imaging data, including high computational complexity, large sample size requirements, and prolonged noise propagation pathways. These factors further constrain their effectiveness on small- to medium-scale datasets. Looking ahead, we believe that 3D neural networks still hold considerable promise in medical imaging applications. By incorporating anatomically informed high-precision automatic segmentation tools, building large-scale multi-center datasets, and designing task-specific lightweight 3D architectures, it may be possible to overcome current limitations and achieve higher prognostic accuracy. Such advancements could significantly improve individualized treatment strategies for gastric cancer and provide technical support for elucidating the role of muscle quality in cancer prognosis.

This study also analyzed the decision basis of the model through attention heatmaps generated by the 2D neural network. The results showed that the heatmaps were primarily focused on the muscle regions surrounding the spine, indicating that muscle imaging information is more valuable than skeletal information in predicting the survival of gastric cancer patients. This finding is consistent with clinical experience, as muscles are more sensitive indicators of a patient's nutritional status compared to bones⁴². The process of generating the attention heatmaps also provides guidance for future optimization, namely by removing irrelevant regions to refine 3D image segmentation and further improve model performance.

Although this study has made significant progress in evaluating the prognosis of gastric cancer patients, it has certain limitations. First, the sample size is relatively small, with the validation set containing only 194 patients from Hanshan, and lacking external validation, which may affect the generalizability of the model. Future studies should incorporate multicenter datasets to improve the model's stability and applicability. Second, the “black-box” nature of deep learning models limits their interpretability in clinical applications⁴³. Although this study used attention heatmaps to reveal part of the decision-making rationale, future efforts should combine traditional statistical methods with deep learning techniques to develop more transparent and interpretable hybrid models. Additionally, variations in imaging equipment and scanning parameters may impact the diagnostic reliability of the model. Future research should explore methods for standardizing imaging data to enhance the robustness

of the model. Fourth, our study builds upon established CNN architectures and conventional classifiers rather than developing novel algorithms. The primary innovation lies in the original integration of multimodal muscle radiomics with clinical parameters to construct a prognostic framework for gastric cancer, which has not been previously explored in this specific context. Future work could benefit from developing more customized network architectures tailored for muscle feature extraction.

Building upon the findings of this study, future research should focus on several key directions to enhance the model's clinical utility and scientific robustness. First, increasing the sample size and incorporating data from multiple centers will be essential for validating the model's generalizability and ensuring its applicability across diverse clinical settings. Second, optimizing deep learning architectures—such as developing hybrid frameworks that leverage the complementary strengths of 2D and 3D networks—may further improve both predictive performance and computational efficiency. Third, the integration of multimodal data sources, including radiomics, genomics, metabolomics, and clinical parameters, holds great potential for improving model accuracy and supporting more comprehensive personalized treatment strategies. To address the issue of interpretability, future studies should also explore the combination of attention-based visualization techniques with conventional statistical analyses, thereby enhancing model transparency and clinical trust. Lastly, advancing the clinical translation of such models will require seamless integration with hospital information systems (HIS) and picture archiving and communication systems (PACS), enabling end-to-end, automated prognostic workflows to support individualized care for gastric cancer patients.

In conclusion, this study combined deep learning and radiomics to develop an innovative model for predicting the five-year postoperative survival rate of gastric cancer patients and validated the critical role of sarcopenia in gastric cancer prognosis assessment. The findings not only provide a scientific basis for the radiological diagnosis of sarcopenia but also offer valuable references for precision medicine in gastric cancer patients.

Conclusion

This study successfully integrated radiomics and deep learning techniques to develop a highly accurate model for predicting five-year survival in gastric cancer patients. By generating attention heatmaps, we further validated the critical role of muscle mass in prognostic assessment. Our findings provide a robust technical foundation for formulating personalized chemotherapy plans and advancing precision medicine in gastric cancer care. Additionally, this study explored the feasibility of using 3D imaging for prognostic prediction, offering insights into future technological directions in gastric cancer research. These results are expected to facilitate the continued optimization and personalization of treatment strategies, ultimately contributing to improved survival outcomes and quality of life for patients.

Data availability

The original patient data are restricted by hospital privacy policies. De-identified imaging features and analyzed datasets are available from the corresponding author upon reasonable request.

Received: 30 March 2025; Accepted: 25 June 2025

Published online: 01 July 2025

References

1. Park, S. E. et al. Loss of skeletal muscle mass during palliative chemotherapy is a poor prognostic factor in patients with advanced gastric cancer. *Sci Rep.* 10(1):17683. Published 2020 Oct 19. (2020). <https://doi.org/10.1038/s41598-020-74765-8>
2. Fukuda, Y. et al. Sarcopenia is associated with severe postoperative complications in elderly gastric cancer patients undergoing gastrectomy. *Gastric Cancer*. **19** (3), 986–993. <https://doi.org/10.1007/s10120-015-0546-4> (2016).
3. Ricciardolo, A. A. et al. Strong impact of sarcopenia as a risk factor of survival in resected gastric cancer patients: first Italian report of a bicentric study. *Updates Surg.* **74** (1), 283–293. <https://doi.org/10.1007/s13304-021-01175-4> (2022).
4. Ozcan, S. G. G. et al. Clinical significance of sarcopenia in patients undergoing treatment for gastric cancer. *Rev. Assoc. Med. Bras.* (1992) **69**(12), e20230568. <https://doi.org/10.1590/1806-9282.20230568> (2023) (Published 2023 Nov 13).
5. Yang, W. et al. Quantifying skeletal muscle wasting during chemoradiotherapy with Jacobian calculations for the prediction of survival and toxicity in patients with gastric cancer. *Eur. J. Surg. Oncol.* **46** (7), 1254–1261. <https://doi.org/10.1016/j.ejso.2020.03.223> (2020).
6. O'Brien, S. et al. Sarcopenia and Post-Operative morbidity and mortality in patients with gastric Cancer. *J. Gastric Cancer*. **18** (3), 242–252. <https://doi.org/10.5230/jgc.2018.18.e25> (2018).
7. Zhuang, C. L. et al. Sarcopenia is an independent predictor of severe postoperative complications and Long-Term survival after radical gastrectomy for gastric cancer: analysis from a Large-Scale cohort. *Med. (Baltim)*. **95** (13), e3164. <https://doi.org/10.1097/MD.0000000000003164> (2016).
8. Cruz-Jentoft, A. J. et al. Sarcopenia: European consensus on definition and diagnosis: report of the European working group on sarcopenia in older people. *Age Ageing*. **39** (4), 412–423. <https://doi.org/10.1093/ageing/afq034> (2010).
9. Kirk, B., Cawthon, P. M. & Cruz-Jentoft, A. J. Global consensus for sarcopenia. *Aging (Albany NY)*. **16** (11), 9306–9308. <https://doi.org/10.18632/aging.205919> (2024).
10. Dent, E. et al. International clinical practice guidelines for sarcopenia (ICFSR): screening, diagnosis and management. *J. Nutr. Health Aging*. **22** (10), 1148–1161. <https://doi.org/10.1007/s12603-018-1139-9> (2018).
11. Yao, J. et al. Prevalence of possible sarcopenia in community-dwelling older Chinese adults: a cross-sectional study. *BMJ Open*. **12** (12), e067425. <https://doi.org/10.1136/bmjopen-2022-067425> (2022). Published 2022 Dec 14.
12. Bertschi, D., Kiss, C. M., Beerli, N., Mauthner, O. & Kressig, R. W. Impact of sarcopenia on daily functioning: a cross-sectional study among older inpatients. *Aging Clin. Exp. Res.* **34** (9), 2041–2046. <https://doi.org/10.1007/s40520-022-02175-z> (2022).
13. Tagliafico, A. S., Bignotti, B., Torri, L. & Rossi, F. Sarcopenia: how to measure, when and why. *Radiol. Med.* **127** (3), 228–237. <https://doi.org/10.1007/s11547-022-01450-3> (2022).
14. Vogele, D. et al. Sarcopenia - Definition, radiological diagnosis, clinical significance. Sarcopenie – Definition, radiologische erfassung, klinische bedeutung. *Rofo* **195** (5), 393–405. <https://doi.org/10.1055/a-1990-0201> (2023).
15. Gu, Y. et al. Comparative study of XGBoost and logistic regression for predicting sarcopenia in postsurgical gastric cancer patients. *Sci. Rep.* **15** (1), 12808. <https://doi.org/10.1038/s41598-025-98075-z> (2025). PMID: 40229548; PMCID: PMC11997166.

16. Pacholek, K. & Sobieszcańska, M. Sarcopenia identification during comprehensive geriatric assessment. *Int. J. Environ. Res. Public Health*. **19** (1), 32. <https://doi.org/10.3390/ijerph19010032> (2021). PMID: 35010295; PMCID: PMC8751172.
17. Dong, X. et al. Identifying sarcopenia in advanced non-small cell lung cancer patients using skeletal muscle CT radiomics and machine learning. *Thorac. Cancer*. **11** (9), 2650–2659. <https://doi.org/10.1111/1759-7714.13598> (2020).
18. Liu, Z. et al. Deep learning-based radiomics allows for a more accurate assessment of sarcopenia as a prognostic factor in hepatocellular carcinoma. *J. Zhejiang Univ. Sci. B*. **25** (1), 83–90. <https://doi.org/10.1631/jzus.B2300363> (2024).
19. Kumar, V. et al. Radiomics: the process and the challenges. *Magn. Reson. Imaging*. **30** (9), 1234–1248. <https://doi.org/10.1016/j.mri.2012.06.010> (2012).
20. Zumsteg, D. M., Chu, C. E. & Midwinter, M. J. Radiographic assessment of sarcopenia in the trauma setting: a systematic review. *Trauma. Surg. Acute Care Open*. **5** (1), e000414. <https://doi.org/10.1136/tsaco-2019-000414> (2020). Published 2020 Mar 15.
21. Shin, J. et al. A radiomics-based model for predicting prognosis of locally advanced gastric cancer in the preoperative setting. *Sci Rep.* **11** (1):1879. Published 2021 Jan 21. (2021). <https://doi.org/10.1038/s41598-021-81408-z>
22. Tomaszewski, M. R. & Gillies, R. J. The Biological Meaning of Radiomic Features [published correction appears in *Radiology*. **299**(2):E256. (2021). <https://doi.org/10.1148/radiol.2021219005>.]. *Radiology*. 2021;298(3):505–516. doi:10.1148/radiol.2021202553.
23. Basirinia, G. et al. Theranostic approaches for gastric cancer: an overview of in vitro and in vivo investigations. *Cancers (Basel)*. **16** (19), 3323. <https://doi.org/10.3390/cancers16193323> (2024). PMID: 39409942; PMCID: PMC11476023.
24. Canfora, I. et al. A predictive system to classify preoperative grading of rectal Cancer using radiomics features. In *Image Analysis and Processing. ICIAP 2022 Workshops. ICIAP 2022 Vol. 13373* (eds Mazzeo, P. L., Frontoni, E., Sclaroff, S., Distant, C. et al.) (Springer, 2022). https://doi.org/10.1007/978-3-031-13321-3_38.
25. Litjens, G. et al. A survey on deep learning in medical image analysis. *Med. Image Anal.* **42**, 60–88. <https://doi.org/10.1016/j.media.2017.07.005> (2017).
26. Mourtzakis, M. et al. A practical and precise approach to quantification of body composition in cancer patients using computed tomography images acquired during routine care. *Appl Physiol Nutr Metab.* **33**(5):997–1006. (2008). <https://doi.org/10.1139/H08-075>. PMID: 18923576.
27. Shen, W. et al. Total body skeletal muscle and adipose tissue volumes: Estimation from a single abdominal cross-sectional image. *J. Appl. Physiol.* (1985). **97** (6), 2333–2338. <https://doi.org/10.1152/jappphysiol.00744.2004> (2004). Epub 2004 Aug 13. PMID: 15310748.
28. He, K., Zhang, X., Ren, S. & Sun, J. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778. (2016). <https://doi.org/10.1109/CVPR.2016.90>
29. Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K. Q. Densely connected convolutional networks. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*. **4700**, 4708. <https://doi.org/10.1109/CVPR.2017.243> (2017).
30. De Raffele, E., Mirarchi, M., Cuicchi, D., Lecce, F. & Cola, B. Evolving role of FDG-PET/CT in prognostic evaluation of resectable gastric cancer. *World J. Gastroenterol.* **23** (38), 6923–6926. <https://doi.org/10.3748/wjg.v23.i38.6923> (2017).
31. Gillies, R. J., Kinahan, P. E. & Hricak, H. Radiomics: images are more than pictures, they are data. *Radiology* **278** (2), 563–577. <https://doi.org/10.1148/radiol.2015151169> (2016).
32. Wang, Y. et al. CT radiomics nomogram for the preoperative prediction of lymph node metastasis in gastric cancer. *Eur. Radiol.* **30** (2), 976–986. <https://doi.org/10.1007/s00330-019-06398-z> (2020).
33. Ma, T. et al. A machine learning based radiomics approach for predicting no. 14v station lymph node metastasis in gastric cancer. *Front. Med. (Lausanne)*. **11**, 1464632. <https://doi.org/10.3389/fmed.2024.1464632> (2024). Published 2024 Oct 18.
34. Cui, Y. et al. A CT-based deep learning radiomics nomogram for predicting the response to neoadjuvant chemotherapy in patients with locally advanced gastric cancer: A multicenter cohort study. *EclinicalMedicine*. **46**:101348. Published 2022 Mar 21. (2022). <https://doi.org/10.1016/j.eclinm.2022.101348>
35. Yao, L. et al. A review of deep learning and radiomics approaches for pancreatic cancer diagnosis from medical imaging. *Curr. Opin. Gastroenterol.* **39** (5), 436–447. <https://doi.org/10.1097/MOG.0000000000000966> (2023).
36. Chen, Q. et al. Transformer-based skeletal muscle deep-learning model for survival prediction in gastric cancer patients after curative resection. *Gastric Cancer*. **28** (4), 684–695 (2025). Epub 2025 Apr 15. PMID: 40232557; PMCID: PMC12174216.
37. Xiang, Y. et al. Radiomics based on machine learning algorithms could predict prognosis and postoperative chemotherapy benefits of patients with gastric cancer: a retrospective cohort study. *J. Gastrointest. Oncol.* **14** (5), 2048–2063. <https://doi.org/10.21037/jgo-23-627> (2023). Epub 2023 Oct 27. PMID: 37969820; PMCID: PMC10643584.
38. Hao, D. et al. SurvivalCNN: A deep learning-based method for gastric cancer survival prediction using radiological imaging data and clinicopathological variables. *Artif. Intell. Med.* **134**, 102424. <https://doi.org/10.1016/j.artmed.2022.102424> (2022). Epub 2022 Oct 19. PMID: 36462894.
39. Fang, C. et al. Metabolic Reprogramming and Risk Stratification of Hepatocellular Carcinoma Studied by Using Gas Chromatography-Mass Spectrometry-Based Metabolomics. *Cancers (Basel)*. **14**(1):231. Published 2022 Jan 4. (2022). <https://doi.org/10.3390/cancers14010231>
40. Pfob, A., Lu, S. C. & Sidey-Gibbons, C. Machine learning in medicine: a practical introduction to techniques for data pre-processing, hyperparameter tuning, and model comparison. *BMC Med. Res. Methodol.* **22** (1), 282. <https://doi.org/10.1186/s12874-022-01758-8> (2022). PMID: 36319956; PMCID: PMC9624048.
41. Hicks, S. A. et al. On evaluation metrics for medical applications of artificial intelligence. *Sci. Rep.* **12** (1), 5979. <https://doi.org/10.1038/s41598-022-09954-8> (2022). PMID: 35395867; PMCID: PMC8993826.
42. Fensham, N. C. et al. Short-Term carbohydrate restriction impairs bone formation at rest and during prolonged exercise to a greater degree than low energy availability. *J. Bone Min. Res.* **37** (10), 1915–1925. <https://doi.org/10.1002/jbmr.4658> (2022).
43. Shao, Y. et al. Shedding Light on the Black Box: Explaining Deep Neural Network Prediction of Clinical Outcomes. *J Med Syst.* **45**(1):5. Published 2021 Jan 4. (2021). <https://doi.org/10.1007/s10916-020-01701-8>

Acknowledgements

We thank the First Affiliated Hospital of Anhui Medical University and the People's Hospital of Henshan County, Anhui Province, for granting authorisation for the use of medical data and ethical approval for this article. We are also grateful to onekey technology platform for its artificial intelligence technology support.

Author contributions

Yuan Hong: Conceptualization, Project administration, Visualization, Writing original draft, Peng Zhang: Data curation, hyperparameter tuning, and model comparison. BMC Med. Res. Methodol. **22** (1), 282. <https://doi.org/10.1186/s12874-022-01758-8> (2022). PMID: 36319956; PMCID: PMC9624048.

Hicks, S. A. et al. On evaluation metrics for medical applications of artificial intelligence. *Sci. Rep.* **12** (1), 5979. <https://doi.org/10.1038/s41598-022-09954-8> (2022). PMID: 35395867; PMCID: PMC8993826.

Fensham, N. C. et al. Short-Term carbohydrate restriction impairs bone formation at rest and during prolonged exercise to a greater degree than low energy availability. *J. Bone Min. Res.* **37** (10), 1915–1925. <https://doi.org/10.1002/jbmr.4658> (2022).

Shao, Y. et al. Shedding Light on the Black Box: Explaining Deep Neural Network Prediction of Clinical Outcomes. *J Med Syst.* **45**(1):5. Published 2021 Jan 4. (2021). <https://doi.org/10.1007/s10916-020-01701-8>

Declarations

Competing interests

The authors declare no potential conflict of interest in the research, writing, and publication of this article.

Financial support

This study was supported by a grant from the Graduate Student Research and Practice Innovation Program of Anhui Medical University (YJS20230080), National Natural Science Foundation of China (82403333).

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-09083-y>.

Correspondence and requests for materials should be addressed to G.C. or B.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025