

Functional Agent-Driven Computational Anatomy for Preoperative Siewert Classification of Adenocarcinoma of the Esophagogastric Junction Using Three-Dimensional Contrast- Enhanced CT: a retrospective multicenter diagnostic accuracy study

Yiting Liu

Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education, Beijing),
Department of Radiology, Peking University Cancer Hospital and Institute

Zifan Chen

Center for Data Science, Peking University

Ruojin Li

School of Information and Intelligent Science, Donghua University

Shuangxi Li

Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Department of Gastrointestinal Cancer Center, Peking University Cancer Hospital and Institute

Li Zhang

Center for Data Science, Peking University

Xiaoting Li

Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education, Beijing),
Department of Radiology, Peking University Cancer Hospital and Institute

Yinkui Wang

Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Department of Gastrointestinal Cancer Center, Peking University Cancer Hospital and Institute

Kan Xue

Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Department of Gastrointestinal Cancer Center, Peking University Cancer Hospital and Institute

Fei Shan

State Key Laboratory of Holistic Integrative Management of Gastrointestinal Cancers, Beijing Key Laboratory of Carcinogenesis and Translational Research, Department of Gastrointestinal Cancer Center

Lei Tang

tanglei@pku.edu.cn

Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education, Beijing),
Department of Radiology, Peking University Cancer Hospital and Institute

Ziyu Li

Key Laboratory of Carcinogenesis and Translational Research (Ministry of
Education/Beijing), Department of Gastrointestinal Cancer Center, Peking University Cancer Hospital and
Institute

Research Article

Keywords: Adenocarcinoma of the esophagogastric junction, Siewert classification, Image
segmentation, Computed Tomography

Posted Date: October 14th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10961407/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Accurate preoperative Siewert classification of adenocarcinoma of the esophagogastric junction (AEG) is critical for surgical planning, but current imaging relies on subjective estimation, leading to poor reproducibility and potential misclassification. We developed and evaluated FACET, a functional agent-driven computational anatomy framework that converts subjective spatial judgment into calibrated, traceable 3D distance measurements between the tumor midpoint and the esophagogastric junction (EGJ) interface on contrast-enhanced CT.

Methods

In this retrospective multi-cohort study (development $n = 70$; internal validation $n = 61$; external validation $n = 45$), FACET integrated automated tumor perception, anatomy-constrained structure refinement, and mucosal-surface-guided 3D path modeling to localize the tumor midpoint and EGJ interface. Siewert types were determined by calibrated geodesic distances along the path. Histopathological Siewert type served as the reference standard.

Results

With the calibration factor locked at $k = 0.7$, FACET achieved accuracies of 71.4%, 82.0%, and 73.3% in the development, internal validation, and external validation cohorts, respectively; the corresponding AUCs were 0.721, 0.927, and 0.770. In the external validation cohort, balanced accuracy was 76.5%. The complete workflow was implemented as an interactive desktop application for three-dimensional visualization, anatomical measurement, and Siewert classification.

Conclusions

FACET transforms subjective imaging interpretation into reproducible 3D anatomical assessment. In the independent external cohort, it demonstrated moderate discrimination and showed potential for cross-institutional application as a traceable decision-support tool for preoperative multidisciplinary planning in AEG patients.

Background

The incidence of adenocarcinoma of the esophagogastric junction (AEG) has shown a significantly increasing trend in both Western and Eastern populations in recent years[1]. The Siewert classification of AEG plays a pivotal role in systematically guiding surgical approach selection, defining the extent of lymph node dissection (such as lower mediastinal lymphadenectomy), and stratifying prognostic risk[2,

3]. As a core categorization scheme widely endorsed by domestic and international clinical guidelines[4], the Siewert classification[5] delineates AEG into types I, II, and III based on the spatial relationship between the tumor epicenter and the esophagogastric junction (EGJ). This classification scheme elucidates clinically relevant patterns associated with distinct anatomical locations and directly informs key perioperative choices, including multidisciplinary treatment strategy and operative approach pathway[6].

However, current imaging-based classification workflows largely rely on subjective endoscopic estimation or retrospective interpretation of axial computed tomography (CT) images. In complex clinical scenarios—such as locally advanced tumors, irregular tumor morphology, extension across the diaphragmatic hiatus, or indistinct tumor-organ boundaries—the traditional anatomical measurement bias and inter-observer variability become highly pronounced, leading to limited diagnostic accuracy and suboptimal reproducibility[7, 8]. Recent studies have also noted that conventional endoscopy frequently miscalculates or underestimates the longitudinal span of submucosal invasion in AEG[9], failing to delineate the accurate volumetric boundaries of the lesion. Moreover, imaging interpretation and manual measurement for these challenging cases are inherently resource-intensive, unavoidably dependent on expert experience, and associated with substantial clinical workload, further impeding the standardization and broad deployment of Siewert classification in routine practice.

To address these technical bottlenecks, the key methodological need is not simply another automated segmentation model, but a clinically interpretable framework capable of translating the anatomical definition of Siewert classification into a computable preoperative imaging task. In this study, we developed FACET, a functional agent-driven computational anatomy framework that organizes AEG classification into coordinated steps of anatomical perception, spatial reasoning, and rule-based decision-making. By embedding clinical anatomical knowledge into the computational workflow, FACET aims to move Siewert classification beyond experience-dependent image interpretation toward a standardized and anatomically grounded assessment paradigm on three-dimensional contrast-enhanced CT.

Methods

Study Population and Patient Selection

A total of 176 patients with histopathologically confirmed AEG were included in the final analysis to develop, optimize, and validate the proposed computational anatomy framework. This diagnostic accuracy study adheres to international reporting guidelines such as STARD-AI[10]. The multicenter dataset comprised three distinct cohorts, and the same prespecified eligibility criteria were applied across all three cohorts. For the development cohort, 85 patients were screened for eligibility; 15 were excluded according to the prespecified exclusion criteria, leaving 70 patients in the final development cohort. These 70 patients comprised 31 from Peking University Cancer Hospital and 39 from Changzhi People's Hospital of Shanxi Province. For the internal validation cohort (Validation Cohort 1), 110

consecutive patients from Peking University Cancer Hospital were screened; 49 were excluded according to the same prespecified criteria, leaving 61 patients for final analysis. For the external validation cohort (Validation Cohort 2), 57 consecutive patients from Changzhi People's Hospital of Shanxi Province were screened; 12 were excluded according to the same prespecified criteria, leaving 45 patients for final analysis. Thus, the final analysis population comprised 176 patients (development $n = 70$; internal validation $n = 61$; external validation $n = 45$). The patient screening, exclusion, and cohort-assignment process is summarized in Fig. 1.

The inclusion criteria, applied consistently across all three cohorts, were defined as follows: (1) histopathologically confirmed AEG based on endoscopic biopsy or surgical resection; (2) availability of preoperative contrast-enhanced abdominal CT performed within 30 days prior to surgery; and (3) complete, verifiable records of operative specimens including the definitive histopathological Siewert classification. The exclusion criteria, likewise applied across all three cohorts, were: (1) administration of neoadjuvant therapy prior to surgical resection. Patients receiving neoadjuvant therapy were excluded to minimize treatment-related anatomical distortion and preserve a reliable correspondence between preoperative CT-derived spatial measurements and postoperative pathological Siewert classification. This criterion may limit applicability of FACET in contemporary multimodal treatment settings and requires future validation in post-neoadjuvant cohorts; (2) absence or inadequate quality of preoperative abdominal CT; (3) CT imaging performed more than 30 days before surgery; (4) incomplete coverage of the EGJ region on the CT scan; and (5) major imaging artifacts obscuring lesion evaluation. This study was approved by the Medical Ethics Committee of Peking University Cancer Hospital (approval No. 2020KT71) and the Medical Ethics Committee of Changzhi People's Hospital (approval No. 2024K003). Given that this investigation was a completely de-identified retrospective imaging study that did not compromise patient privacy or alter clinical risk, the requirement for informed consent for this retrospective analysis was waived by the relevant ethics committees. Pathological Siewert classification was determined from the surgical specimen according to the conventional position of the tumor epicenter relative to the anatomical esophagogastric junction (EGJ). For specimen assessment, the EGJ was defined morphologically as the transition between the tubular distal esophagus and the proximal stomach/cardial region rather than by the mucosal Z-line alone. The tumor epicenter was determined from the longitudinal extent of the gross lesion, and Siewert types I, II, and III were assigned using the conventional distance boundaries relative to the EGJ. Gross and histopathological findings were integrated when the junction or tumor border was ambiguous. The same anatomical definition and Siewert distance criteria were applied at both participating institutions to improve inter-institutional consistency.

FACET: A Functional Agent-Driven Computational Anatomy Framework for Siewert Classification of AEG

In this study, we present the Functional Agent-driven Computational Anatomy EGJ Tool (FACET), a multi-agent framework for anatomically grounded non-invasive Siewert classification of AEG from three-

dimensional (3D) CT scans. Unlike conventional end-to-end classification approaches [11], FACET reformulates the Siewert classification problem into a sequence of anatomically grounded and clinically measurable sub-tasks, rather than a direct black-box prediction, thereby enabling step-wise reasoning and explicit validation at each step. These tasks combine automated computation with radiologist-provided guidance. The FACET framework integrates a modular system of functional agents, each responsible for a dedicated anatomical task within the classification workflow. These agents are orchestrated to preserve anatomical continuity and spatial consistency across successive stages, enabling step-wise reasoning and explicit verification of intermediate results.

The methodological contribution of FACET does not arise from nnU-Net segmentation or breadth-first search as standalone algorithms. Rather, the innovation lies in translating the clinical definition of Siewert classification into an explicit computational anatomy workflow in which anatomical perception, anatomy-constrained structure selection, mucosal-surface geodesic path construction, hierarchical EGJ interface reasoning, and rule-based distance classification are linked into a traceable sequence. Each intermediate anatomical object can be inspected and, when necessary, corrected by the clinician, distinguishing the framework from direct image-to-label pipelines and from conventional segmentation-plus-measurement workflows that do not explicitly encode the EGJ-centered clinical decision rule.

As shown in Fig. 2[12], the framework begins with the perception agent, which identifies candidate gastric and esophageal tumor regions from CT images. The extracted candidates are subsequently processed by the structure selection agent, which refines and selects anatomically plausible components for downstream analysis, thereby filtering out segmentation noise and resolving ambiguities in complex tumor presentations. The path modeling agent then constructs a mucosal-surface-guided search space, generates the continuous three-dimensional tumor main-axis path, and determines the tumor midpoint. This path provides a quantitative representation of longitudinal tumor extent along the gastric wall and reflects the anatomical orientation considered during endoscopic and surgical evaluation. Subsequently, the interface reasoning agent localizes a representative EGJ interface point through a hierarchical reasoning strategy integrating structural overlap, spatial relationship analysis, and lumen-based anatomical cues.

These complementary strategies are applied sequentially according to the reliability of available anatomical information, thereby maintaining robustness across variable tumor morphologies and imaging conditions. Finally, the geodesic distance along the main-axis path between the tumor midpoint and the inferred interface point is computed and calibrated to derive the Siewert classification. By quantifying the tumor's anatomical features, the FACET system provides a clear and measurable basis for Siewert classification, aiding surgeons in making informed preoperative decisions.

Perception of Candidate Anatomical Structures

The perception agent, as a functional unit within FACET, serves as the initial stage dedicated to comprehensive candidate generation and establishes the foundational anatomical context by extracting

candidate structures critical for Siewert classification. This stage targets two primary anatomical entities: the gastric tumor and the esophageal tumor. The inputs include contrast-enhanced venous-phase 3D CT scans, which are processed to generate candidate anatomical structures for subsequent reasoning and path modeling.

Candidate structures are generated using two anatomy-specific nnU-Net models [11] for gastric and esophageal tumor segmentation, respectively, trained on manually annotated clinical CT datasets (Supplementary Text 1). Importantly, this stage prioritizes sensitivity over specificity, aiming to capture all potentially relevant anatomical components without imposing strict anatomical constraints, which are deferred to the subsequent structure selection stage. During dataset preparation, expert radiologists provided high-quality annotations on representative cases, following standardized delineation protocols, ensuring accurate delineation of tumor regions and exclusion of intraluminal gas, fluid, blood vessels, and imaging artifacts. All annotations were further subjected to multi-planar cross-validation (axial, coronal, and sagittal views) to ensure volumetric consistency and boundary accuracy. Once trained, the models perform fully automated segmentation of new CT scans, producing 3D candidate volumes for downstream structure selection and path modeling. These candidate volumes can be directly visualized and inspected, enabling optional clinician verification in cases with complex anatomy or ambiguous tumor boundaries.

By preserving the spatial attributes of each candidate structure, the perception agent provides the anatomical basis for subsequent structure selection, path modeling, and interface reasoning within FACET.

Structure Selection and Path Modeling

The structure selection agent and path modeling agent, as functional units within FACET, refine candidate structures and construct the tumor main-axis path, forming the quantitative basis for Siewert classification. In contrast to the perception stage, which prioritizes comprehensive candidate generation, this stage explicitly introduces anatomical constraints to ensure structural plausibility and measurement consistency.

Following the initial perception stage, candidate anatomical structures are first processed by the structure selection agent, which applies anatomical constraints to eliminate disconnected or spatially implausible components and to resolve multiple candidate regions generated during automated segmentation. This module performs anatomy-guided structure selection based on spatial continuity and anatomical coherence. A clinician-provided tumor starting point is used as a spatial anchor to identify the proximal tumor region and guide subsequent path exploration. Candidate tumor structures are further evaluated according to their spatial relationship with the anchor region, ensuring selection of anatomically plausible structures for downstream analysis.

Using the refined candidates, the path modeling agent constructs a mucosal-surface-constrained search space along the gastric wall, enabling exploration of the tumor's spatial extent. This constraint restricts the search to anatomically feasible regions, preventing shortcuts through non-physiological paths (e.g., across lumen or outside the organ boundary). A three-dimensional breadth-first search (BFS) algorithm is performed within this search space, starting from a spatial anchor provided by the clinician, to generate the continuous tumor main-axis path (Supplementary Fig. 1). The system generates a clear path that mirrors the tumor's progression along the gastric wall, providing surgeons with an easily interpretable guide for planning surgical intervention or evaluating tumor resection. The cumulative geodesic path length along this trajectory is computed, and the point situated at half of this length is defined as the tumor midpoint, which serves as a spatial reference for downstream interface reasoning. Defining the midpoint along the geodesic path, rather than in Euclidean space, ensures consistency with curved anatomical structures and improves robustness under irregular tumor morphology.

The resulting tumor main-axis path provides the spatial foundation for subsequent EGJ localization and path-level distance measurement.

Interface Reasoning and Siewert Classification

The interface reasoning agents and classification module are responsible for localizing the EGJ interface point and determining the Siewert classification based on path-level measurements. Using the refined candidate structures and the tumor main-axis path generated in the preceding stage, the interface reasoning agent localizes a representative EGJ interface point corresponding to the anatomical transition between the esophageal and gastric compartments along the tumor trajectory, while accounting for variable tumor morphology, imaging artifacts, and segmentation uncertainties.

On CT, FACET defines the EGJ anatomically as the transition between the tubular distal esophagus and the gastric cardia/fundus, rather than by the endoscopic squamocolumnar junction (Z-line). This distinction is important in Barrett's esophagus, in which proximal displacement of the squamocolumnar junction may cause the Z-line to diverge from the anatomical EGJ. Accordingly, the interface reasoning agent relies on structural CT relationships among the distal esophagus, proximal stomach, lumen continuity, and the tumor trajectory rather than on an epithelial landmark. In cases with marked distortion or equivocal anatomy, the inferred interface point is intended to undergo clinician verification before final classification [29].

The reasoning process integrates multiple complementary strategies that are organized as an evidence-driven decision hierarchy and automatically orchestrated according to the availability of structural evidence (Supplementary Fig. 2):

Initially, an intersection-based approach evaluates the spatial overlap between the gastric and esophageal tumors along the main-axis path. Limited morphological dilation is applied when necessary to compensate for minor segmentation discrepancies. When both anatomical compartments are

simultaneously represented and the overlap region can be directly associated with the main-axis path, this strategy provides a direct and high-confidence estimate of the EGJ interface point.

When overlap evidence is available but cannot be directly linked to the main-axis path because of local spatial displacement, a projection-based strategy is activated, in which the nearest path location to the candidate overlap region is selected as the EGJ interface point. This strategy preserves anatomical correspondence while ensuring that the final interface point remains located on the main-axis path.

In cases where no reliable overlap evidence exists between the gastric tumor and esophageal tumor, the reasoning agent invokes a distance-balancing strategy. This strategy evaluates the shortest distances from each path location to the gastric and esophageal tumors and identifies the location exhibiting the smallest difference between the two distances, thereby representing the point with the most balanced spatial relationship to both anatomical compartments and approximating the anatomical transition zone at the EGJ.

Finally, when either tumor is unavailable and structural evidence becomes insufficient for direct interface reasoning, a lumen-density-based fallback strategy is employed. Local variations in gastric lumen density along the tumor trajectory are analyzed to infer the most probable EGJ location based on characteristic anatomical changes occurring across the esophagogastric junction.

Through this evidence-driven orchestration process, the interface reasoning agent adaptively integrates structural consistency, spatial correspondence, anatomical proximity, and lumen-based contextual information, enabling robust EGJ localization across diverse tumor morphologies and varying segmentation quality.

Following EGJ interface localization, the classification module computes the geodesic distance (d) along the main-axis path between the tumor midpoint and the inferred interface point. The calibrated distance was defined as $d' = k \times d$, where the sign indicates whether the tumor midpoint lies on the esophageal or gastric side of the interface. Candidate values of k were evaluated in the development cohort using classification accuracy as the optimization criterion. The resulting calibration factor was fixed at $k = 0.7$ before evaluation in the internal and external validation cohorts and was not subsequently adjusted. Using the conventional calibrated-distance boundaries, $d' < -1.0$ cm, $-1.0 \leq d' \leq 2.0$ cm, and $d' > 2.0$ cm corresponded to Siewert types I, II, and III, respectively. For the prespecified binary analysis, $d' \leq 2.0$ cm was classified as Siewert type I/II and $d' > 2.0$ cm as type III; at $k = 0.7$, the 2.0-cm calibrated boundary corresponds to $d = 2.86$ cm on the raw distance scale.

Clinical Software Implementation

To facilitate clinical implementation of the proposed FACET framework, we developed a dedicated CT-based AEG analysis software platform (Fig. 3). The platform integrates volumetric CT visualization, automated tumor segmentation, anatomy-constrained tumor path reconstruction, EGJ interface localization, Siewert classification, three-dimensional rendering, and standardized result export within a

unified workflow. Intermediate outputs generated by each functional agent, including tumor masks, the tumor main-axis path, the path-defined tumor midpoint, and the inferred EGJ interface point, can be displayed and interactively reviewed during the analysis process. For cases with complex anatomy, segmentation uncertainty, or ambiguous tumor boundaries, the platform allows optional manual mask refinement and clinician verification, thereby supporting anatomical plausibility, measurement reliability, and traceability of the final classification.

Statistical Analysis

Statistical analyses were performed using SPSS version 26.0 (IBM, Armonk, NY, USA). Continuous variables were summarized as the median and interquartile range (IQR), and categorical variables were presented as counts and percentages. Continuous variables were compared using the Mann–Whitney U test, whereas categorical variables were compared using the chi-square test or Fisher’s exact test, as appropriate. Inter-reader agreement was assessed using Cohen’s kappa (κ) statistic. A κ value of ≥ 0.75 was considered excellent agreement, 0.40 to < 0.75 moderate-to-good agreement, and ≤ 0.40 poor agreement. Histopathological Siewert classification was used as the reference standard. For the prespecified binary analysis, Siewert types I/II were defined as the positive class and type III as the negative class. Accuracy, sensitivity, specificity, PPV, and NPV were reported with 95% CIs estimated using the Wilson score method. F1 score and balanced accuracy were evaluated at the same locked operating point, with 95% CIs obtained using percentile bootstrap resampling with 2,000 replications. ROC curves were constructed using the negative continuous raw distance (-d) as the discrimination score, and AUCs with 95% CIs were estimated using percentile bootstrap resampling with 2,000 replications. Differences in the distribution of Siewert types I/II and III across cohorts were assessed using Pearson’s chi-square test. All tests were two-sided, and $P < 0.05$ was considered statistically significant. Because FACET outputs a continuous anatomical distance and a rule-based Siewert class rather than an estimated event probability, conventional probability-calibration plots were not used; performance evaluation therefore focused on discrimination and threshold-based diagnostic accuracy. Decision-curve analysis was not performed because the present study did not define a probability-based treatment threshold or a treatment-triggering risk model.

Use of artificial intelligence tools

ChatGPT (OpenAI) was used solely to assist with manuscript language editing and formatting during manuscript preparation. All scientific content, analyses, interpretations, and final wording were reviewed and approved by the authors. No artificial intelligence tool was credited as an author.

Results

Patient distributions based on the final diagnostic-performance analysis set are summarized in Table 1. The analysis included 176 patients: 70 in the development cohort, 61 in the internal validation cohort,

and 45 in the external validation cohort. Siewert classifications I/II accounted for 99 of 176 patients (56.3%), and Siewert classification III accounted for 77 of 176 patients (43.8%). The Siewert distribution differed significantly across cohorts ($P < 0.001$).

Table 1
Patient Distribution Across Final Analysis Cohorts

Characteristic	Development cohort-(n = 70)	Internal validation(n = 61)	External validation(n = 45)	Overall(n = 176)	P value
Final analysis set, n	70	61	45	176	
Siewert classification, n (%)					< 0.001
I/II, n (%)	32 (45.7)	34 (55.7)	33 (73.3)	99 (56.3)	
III, n (%)	38 (54.3)	27 (44.3)	12 (26.7)	77 (43.8)	
Note: Table 1 summarizes the final diagnostic-performance analysis set (n = 176), consistent with the Methods and Table 2. Percentages were calculated using the corresponding cohort denominators. Siewert classifications I and II were grouped as I/II for consistency with the binary diagnostic-performance analysis; this grouping was used solely for model-performance statistics and does not alter the conventional clinicopathological Siewert classification.					

The diagnostic performance of FACET was evaluated in the development cohort and the internal and external validation cohorts, using histopathological Siewert classification as the reference standard. The locked calibration factor ($k = 0.7$) was applied consistently across all three cohorts. The complete performance metrics, including accuracy, sensitivity, specificity, positive predictive value, negative predictive value, F1 score, balanced accuracy, and area under the receiver operating characteristic curve (AUC), together with their 95% confidence intervals, are summarized in Table 2. Receiver operating characteristic curves are presented in Fig. 4.

In the development cohort (n = 70), FACET achieved an accuracy of 71.4%, a sensitivity of 71.9%, a specificity of 71.1%, and an AUC of 0.721. In the internal validation cohort (n = 61), the corresponding values were 82.0%, 82.4%, 81.5%, and 0.927, respectively. In the external validation cohort (n = 45), FACET achieved an accuracy of 73.3%, a sensitivity of 69.7%, a specificity of 83.3%, and an AUC of 0.770. Detailed performance estimates, including 95% confidence intervals, are provided in Table 2.

Table 2
Diagnostic Performance Metrics and Siewert Tumor Distributions Across Development and Validation Cohorts

Metrics / Inputs	Development Cohort-	Internal Validation Cohort	External Validation Cohort
True Positive (TP)	23	28	23
True Negative (TN)	27	22	10
False Positive (FP)	11	5	2
False Negative (FN)	9	6	10
Total (n)	70	61	45
Accuracy, % (95% CI)	71.4 (59.9–80.7)	82.0 (70.5–89.6)	73.3 (59.0–84.0)
Sensitivity, % (95% CI)	71.9 (54.6–84.4)	82.4 (66.5–91.7)	69.7 (52.7–82.6)
Specificity, % (95% CI)	71.1 (55.2–83.0)	81.5 (63.3–91.8)	83.3 (55.2–95.3)
PPV, % (95% CI)	67.6 (50.8–80.9)	84.8 (69.1–93.3)	92.0 (75.0–97.8)
NPV, % (95% CI)	75.0 (58.9–86.2)	78.6 (60.5–89.8)	50.0 (29.9–70.1)
F1 score (95% CI)	0.697 (0.556–0.814)	0.836 (0.724–0.923)	0.793 (0.653–0.897)
Balanced accuracy, % (95% CI)	71.5 (60.1–81.9)	81.9 (71.9–91.4)	76.5 (62.5–88.5)
AUC (95% CI)	0.721 (0.595–0.839)	0.927 (0.856–0.978)	0.770 (0.606–0.909)
Note: Data are presented as counts, percentages with 95% confidence intervals (CIs), or estimates with 95% CIs. Histopathological Siewert classification was used as the reference standard. Siewert types I/II were prespecified as the positive class, and Siewert type III as the negative class. The calibration factor $k = 0.7$, selected in the development cohort, was locked and applied unchanged to the internal and external validation cohorts. The negative continuous calibrated distance ($-d'$) was used as the score for receiver operating characteristic analysis, whereas threshold-based performance metrics were calculated using the predefined 2-cm classification boundary. AUC = area under the receiver operating characteristic curve; CI = confidence interval; FN = false negative; FP = false positive; NPV = negative predictive value; PPV = positive predictive value; TN = true negative; TP = true positive.			

Discussion

The present study demonstrates that FACET, a functional agent-driven computational anatomy framework, provides a reliable and non-invasive method for preoperative Siewert classification of AEG using 3D contrast-enhanced CT. Beyond a conventional segmentation or image-label classification model, FACET reframes Siewert classification as a computational anatomy problem: the clinical rule based on the spatial relationship between the tumor epicenter and the EGJ is translated into a structured, measurable, and auditable imaging workflow. This design is particularly relevant for AEG,

where tumor morphology, mucosal extension, and the EGJ interface are often difficult to interpret consistently on routine axial images.

Accurate tumor boundary localization and spatial delineation are fundamental for determining the optimal surgical approach, defining the required extent of gastric/esophageal resections, and planning lymph node dissection strategies. Securing an adequate proximal resection margin (PM) is clinically important, because microscopic margin involvement (R1 status) is a critical independent predictor of local recurrence and poor survival outcomes[13]. Multiple retrospective studies have highlighted that the exact length of the negative proximal margin correlates with localized disease control and long-term overall survival in AEG patients[14, 15], especially when transabdominal or transdiaphragmatic approaches are considered. Therefore, a method that measures tumor extent along the anatomically curved mucosal surface is conceptually closer to operative and pathological assessment than conventional axial or coronal estimations.

A central strength of FACET is its integration of data-driven anatomical perception with anatomy-driven spatial reasoning. Automated segmentation methods, including nnU-Net-based models, have shown favorable performance in several tumor-related tasks, such as venous-phase CT pancreas segmentation (Dice = 0.805)[16], MRI-based oropharyngeal squamous cell carcinoma delineation (Dice = 0.77)[17], and automated gastric cancer segmentation for Lauren classification prediction (Dice = 0.79)[18]. In the present framework, segmentation is not treated as the final classification output, but as an upstream anatomical perception step that supports subsequent structure refinement, path modeling, and interface reasoning. This organization helps preserve interpretability and avoids reducing Siewert classification to a direct black-box prediction problem.

The segmentation-by-parts strategy used in FACET is also anatomically motivated. Compared with gastric lesions, fully automated esophageal tumor segmentation on CT remains challenging because of poor esophageal distension, respiratory and swallowing-related motion artifacts near the diaphragmatic hiatus, irregular tumor morphology, and indistinct tumor-wall boundaries on contrast-enhanced CT. These factors are consistent with previous reports showing lower segmentation performance for esophageal cancer than for structurally more regular tumors. Accordingly, FACET processes the cross-anatomical components of AEG separately: the esophageal component is handled with a clinician-anchored semi-automated length-tracking strategy to preserve precision, whereas the gastric component is segmented automatically where volumetric modeling has a more mature technical foundation. This partitioned design reduces the impact of morphological heterogeneity and supports more stable boundary definition across centers.

The 3D BFS component further extends the framework from anatomical perception to spatial measurement. As a graph traversal strategy based on layer-by-layer neighborhood expansion, 3D BFS is well suited for estimating geodesic distances over curved surfaces such as the gastric mucosal surface. In this task, the major technical challenge is not simply finding the longest diameter, but ensuring that the measured trajectory follows the longitudinal direction of the gastric lumen rather than an

inappropriate circumferential route across the greater and lesser curvatures or between the anterior and posterior walls. By restricting traversal to the mucosal VOI and defining bounded start and end regions, FACET enables ordered point traversal on the curved surface, reduces redundant computation, and makes the resulting path more consistent with surgical and pathological length measurement.

This graph-based formulation also addresses a known limitation of search algorithms: sensitivity to initial point placement. In FACET, the path computation is supported by anatomical constraints and automatically extracted mucosal-surface features, while the clinician-provided anchor is used to maintain clinical plausibility in anatomically ambiguous regions. In this way, the framework combines human anatomical knowledge with algorithmic path optimization, improving the repeatability of tumor midpoint localization and reducing dependence on purely subjective manual measurement.

Furthermore, a methodological feature of FACET is the incorporation of a linear scale calibration factor k to reconcile CT-derived in vivo measurements with the specimen-based anatomical boundaries used for pathological Siewert classification. In the present study, k was empirically selected in the development cohort by optimizing classification accuracy and was fixed at 0.7 before evaluation in the internal and external validation cohorts. Notably, this development-derived value is close to previously reported ratios between in vivo and specimen-based esophageal measurements [19–21]. Pang et al. reported that a 2.0 cm proximal margin measured on a fixed pathological specimen corresponded to approximately 2.8 cm in vivo, equivalent to a ratio of approximately 0.71 [19], while Shin et al. observed further shrinkage after formalin fixation, to approximately 82% of the fresh ex vivo length [20]. These findings provide biological plausibility for the direction and approximate magnitude of the calibration used in FACET. Nevertheless, $k = 0.7$ should not be interpreted as a universal physiological shrinkage coefficient, because tissue deformation may vary across patients, surgical procedures, specimen-handling protocols, and anatomical measurement paths. Rather, it represents a development-derived calibration parameter whose generalizability requires further prospective validation.

An important limitation of this strategy is that the index test and the reference standard are obtained in different physical states: FACET measures the EGJ-to-tumor relationship in vivo on CT, whereas pathological Siewert classification is determined ex vivo after resection and may be affected by tissue relaxation, traction, specimen orientation, fixation, and handling. The linear calibration factor can reduce a systematic scale difference between CT-derived and specimen-derived distances, but it cannot fully correct patient-specific or direction-dependent deformation. Histopathological Siewert classification should therefore be regarded as a clinically accepted but imperfect reference standard rather than an error-free ground truth. Future prospective studies should compare FACET with standardized intraoperative or fresh-specimen measurements obtained before fixation and should assess whether multimodal consensus localization of the EGJ further reduces reference-standard bias.

Currently, clinical Siewert classification primarily relies on traditional diagnostic modalities, including X-ray barium studies, conventional CT MPR, and EUS. However, these approaches remain highly subjective and may be error-prone for large, irregular, or morphologically advanced tumors[7, 8]. By mapping

distances directly on the segmented mucosal VOI model, FACET converts experience-dependent spatial interpretation into a reproducible and traceable 3D measurement. FACET achieved accuracies of 71.4%, 82.0%, and 73.3%, with corresponding AUCs of 0.721, 0.927, and 0.770, in the development, internal validation, and external validation cohorts, respectively. Given the recognized challenge of dataset shift in clinical artificial intelligence [22, 23], the moderate discrimination retained in the independent external cohort provides initial support for the cross-institutional feasibility of this anatomy-guided approach. FACET should nevertheless be regarded as an interpretable clinical decision-support tool rather than a replacement for expert assessment. Accordingly, we implemented the complete workflow as an interactive desktop application integrating three-dimensional visualization, anatomical measurement, automated classification, and clinician review, allowing clinicians to inspect the anatomical basis and measurement process underlying each classification result.

The distribution of pathological Siewert types differed among the three cohorts (Table 1), consistent with previous clinical series showing variation across institutions, study periods, staging practices, referral patterns, and treatment pathways [24–28]. In particular, the external cohort contained a higher proportion of type I/II tumors than the development and internal validation cohorts. This distribution contributed to the high external PPV of 92.0% and the comparatively low NPV of 50.0%, because both predictive values depend on outcome prevalence. Sensitivity and specificity are less directly affected by prevalence; nevertheless, the external sensitivity of 69.7% indicates that approximately 30% of pathological type I/II tumors were misclassified as type III. Conversely, the specificity of 83.3% was based on only 12 type III cases and therefore had considerable statistical uncertainty. These findings indicate that the high PPV should not be interpreted in isolation as evidence of strong generalizability. Instead, the external AUC of 0.770, balanced accuracy of 76.5%, and the complete set of threshold-based metrics together suggest moderate discriminatory performance while also identifying clinically relevant room for improvement.

Several limitations should be acknowledged. First, this retrospective multi-cohort study remains susceptible to selection bias and institutional differences in referral and surgical selection. During screening, 85 patients were assessed for the development cohort and 15 were excluded, 110 were assessed for the internal validation cohort and 49 were excluded, and 57 were assessed for the external validation cohort and 12 were excluded, all according to the same prespecified eligibility criteria. Because the exclusion criteria included prior neoadjuvant therapy, absence or inadequate quality of preoperative CT, CT performed more than 30 days before surgery, incomplete EGJ coverage, and major imaging artifacts, the evaluated population preferentially represents patients with interpretable and appropriately timed preoperative CT examinations. The available screening records report the total number excluded in each cohort but not the number attributable to each individual exclusion category; therefore, the relative contribution of specific exclusion mechanisms to selection bias cannot be quantified. Second, cases with both eligible preoperative CT imaging and definitive postoperative pathological Siewert classification are inherently scarce, and the validation sample sizes remained modest, particularly the 12 type III cases in the external cohort. Consequently, the specificity estimate remained imprecise, and larger prospective multicenter cohorts with more balanced representation of

Siewert categories are required. Third, FACET has not been specifically validated in patients with large hiatal hernias or other marked anatomical distortions around the EGJ. Herniation or displacement of the gastric cardia may alter the expected relationship among the distal esophagus, diaphragmatic hiatus, proximal stomach, and mucosal path, potentially affecting EGJ localization and geodesic distance measurement. In such cases, clinician review of intermediate anatomical outputs should be considered essential, and dedicated validation is required. Fourth, patients who received neoadjuvant therapy were excluded because treatment-related regression, fibrosis, edema, and remodeling can alter both tumor boundaries and the EGJ-to-tumor spatial relationship. Although this exclusion improved comparability between preoperative CT measurements and postoperative pathology in the present study, it limits applicability in contemporary multimodal treatment settings; FACET therefore requires separate validation in post-neoadjuvant cohorts before such use can be recommended. Fifth, the framework remains partially dependent on the initial automated segmentation quality and the accuracy of the clinician-provided tumor starting point. Sixth, while pathological Siewert classification served as the reference standard, specimen-based localization is subject to post-resection deformation and does not perfectly reproduce the in vivo CT geometry. Finally, the present study evaluates classification accuracy rather than downstream clinical outcomes. FACET produces a deterministic anatomical distance and rule-based class rather than a calibrated probability estimate; consequently, conventional probability calibration and decision-curve analysis were not included. Whether FACET shortens multidisciplinary reporting times, reduces inter-observer variability, improves treatment selection, or yields net clinical benefit should be assessed prospectively.

Conclusions

In conclusion, FACET offers an interpretable, anatomy-guided, and quantitatively traceable framework for CT-based Siewert classification of AEG. By integrating anatomical perception, structure refinement, mucosal-surface-constrained path modeling, and EGJ interface reasoning, the framework translates subjective spatial judgment into a reproducible computational anatomy assessment. The independent external cohort showed moderate discrimination, supporting the feasibility of further multicenter evaluation while also indicating room for improvement before broader clinical adoption.

Abbreviations

AEG
adenocarcinoma of the esophagogastric junction
AUC
area under the receiver operating characteristic curve
BFS
breadth-first search
CI
confidence interval
CT

computed tomography
EGJ
esophagogastric junction
EUS
endoscopic ultrasonography
FACET
Functional Agent-driven Computational Anatomy EGJ Tool
FN
false negative
FP
false positive
IQR
interquartile range
MPR
multiplanar reconstruction
NPV
negative predictive value
PM
proximal resection margin
PPV
positive predictive value
ROC
receiver operating characteristic
STARD-AI
Standards for Reporting Diagnostic Accuracy Studies–Artificial Intelligence
TN
true negative
TP
true positive
VOI
volume of interest.

Declarations

Ethics approval and consent to participate

This retrospective multicenter imaging study was approved by the Medical Ethics Committee of Peking University Cancer Hospital (approval No. 2020KT71) and the Medical Ethics Committee of Changzhi People's Hospital (approval No. 2024K003). The requirement for informed consent for this retrospective analysis was waived by the relevant ethics committees because the study used completely de-identified

clinical imaging data and did not alter patient management. All methods were performed in accordance with the Declaration of Helsinki and relevant institutional guidelines.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and/or analysed during the current study are not publicly available because they contain patient-level clinical imaging data, but are available from the corresponding author on reasonable request and with approval from the institutional ethics committees.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Noncommunicable Chronic Diseases-National Science and Technology Major Project (2024ZD0520600) and the National Natural Science Foundation of China (No. U24A20759, 82441018, and 92459302). The funding bodies had no role in the study design; collection, analysis, and interpretation of data; writing of the report; or the decision to submit the article for publication.

Authors' contributions

YL, LT, and ZL conceptualized the study. YL, ZC, LT, and ZL developed the methodology. RL and ZC developed the software. YL, RL, SL, LZ, XL, YW, KX, and FS performed data curation and investigation. YL, RL, SL, LZ, XL, YW, KX, and FS performed the formal analysis. YL, SL, LT, and ZL interpreted the results. YL and RL drafted the manuscript. LT, ZL, and all authors reviewed and edited the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors thank the study participants and all investigators and clinical staff involved in data collection and management.

References

1. Imamura Y, Watanabe M, Toihata T, et al. Recent incidence trend of surgically resected esophagogastric junction adenocarcinoma and microsatellite instability status in Japanese patients. *Digestion*. 2018;99(1):6–13.
2. Kurokawa Y, Takeuchi H, Doki Y, et al. Mapping of lymph node metastasis from esophagogastric junction tumors: a prospective nationwide multicenter study. *Ann Surg*. 2021;274(1):120–7.

3. Yamashita H, Seto Y, Sano T, et al. Results of a nation-wide retrospective study of lymphadenectomy for esophagogastric junction carcinoma. *Gastric Cancer*. 2017;20:69–83.
4. Obermannová R, Alsina M, Cervantes A, et al. Oesophageal cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. *Ann Oncol*. 2022;33(10):992–1004.
5. Siewert JR, Stein HJ. Classification of adenocarcinoma of the oesophagogastric junction. *Br J Surg*. 1998;85(11):1457–9.
6. Seo SH. Surgical treatment of gastroesophageal junction cancer. *Foregut Surg*. 2023;3(2):21–6.
7. Pedrazzani C, Bernini M, Giacomuzzi S, et al. Evaluation of Siewert classification in gastroesophageal junction adenocarcinoma: what is the role of endoscopic ultrasonography? *J Surg Oncol*. 2005;91(4):226–31.
8. Curtis NJ, Noble F, Bailey IS, et al. The relevance of the Siewert classification in the era of multimodal therapy for adenocarcinoma of the gastro-oesophageal junction. *J Surg Oncol*. 2014;109(3):202–7.
9. Tomi Y, Yamashita K, Umehara M, et al. Pathological invasion length discrepancies and factors influencing endoscopic invasion length assessment in esophagogastric junction adenocarcinoma. *Surg Today*. 2023;53(5):456–64.
10. Sounderajah V, Guni A, Liu X, et al. The STARD-AI reporting guideline for diagnostic accuracy studies using artificial intelligence. *Nat Med*. 2025;31:3283–9.
11. Isensee F, Jaeger PF, Kohl SAA, et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. 2021;18:203–11.
12. Jiang S, Li H, Zhang L, et al. Generic Diagramming Platform (GDP): a comprehensive database of high-quality biomedical graphics. *Nucleic Acids Res*. 2025;53(D1):D1670–6. 10.1093/nar/gkae973.
13. Papachristou DN, Agnanti N, D'Agostino H, et al. Histologically positive esophageal margin in the surgical treatment of gastric cancer. *Am J Surg*. 1980;139(5):711–3.
14. Barbour AP, Rizk NP, Gonen M, et al. Adenocarcinoma of the gastroesophageal junction: influence of esophageal resection margin and operative approach on outcome. *Ann Surg*. 2007;246(1):1–8.
15. Mine S, Sano T, Hiki N, et al. Proximal margin length with transhiatal gastrectomy for Siewert classification II and III adenocarcinomas of the oesophagogastric junction. *Br J Surg*. 2013;100(8):1050–4.
16. Yang E, Kim JH, Min JH, et al. nnU-Net-based pancreas segmentation and volume measurement on CT imaging in patients with pancreatic cancer. *Acad Radiol*. 2024;31(7):2784–94.
17. Choi Y, Bang J, Kim SY, et al. Deep learning-based multimodal segmentation of oropharyngeal squamous cell carcinoma on CT and MRI using self-configuring nnU-Net. *Eur Radiol*. 2024;34(8):5389–400.
18. Cao B, Hu J, Li H, et al. Preoperative prediction of the Lauren classification in gastric cancer using automated nnU-Net and radiomics: a multicenter study. *Insights Imaging*. 2025;16(1):48.

19. Pang T, Nie M, Yin K. The correlation between the margin of resection and prognosis in esophagogastric junction adenocarcinoma. *World J Surg Oncol*. 2023;21:316.
20. Shin Y, Yamamoto M, Ota M, et al. Appropriate proximal resection margin distance for esophagogastric junction adenocarcinoma. *Ann Surg Oncol*. 2026;33:4232–41.
21. Siu KF, Cheung HC, Wong J. Shrinkage of the esophagus after resection for carcinoma. *Ann Surg*. 1986;203(2):173–6.
22. Vasey B, Nagendran M, Campbell B, et al. Reporting guideline for the early-stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *Nat Med*. 2022;28:924–33.
23. Finlayson SG, Subbaswamy A, Singh K, et al. The clinician and dataset shift in artificial intelligence. *N Engl J Med*. 2021;385(3):283–6.
24. Siewert JR, Feith M, Werner M, Stein HJ. Adenocarcinoma of the esophagogastric junction: results of surgical therapy based on anatomical/topographic classification in 1,002 consecutive patients. *Ann Surg*. 2000;232(3):353–61.
25. Feith M, Stein HJ, Siewert JR. Adenocarcinoma of the esophagogastric junction: surgical therapy based on 1602 consecutive resected patients. *Surg Oncol Clin N Am*. 2006;15(4):751–64.
26. Grotenhuis BA, Wijnhoven BPL, Poley JW, et al. Preoperative assessment of tumor location and station-specific lymph node status in patients with adenocarcinoma of the gastroesophageal junction. *World J Surg*. 2013;37(1):147–55.
27. Parry K, Haverkamp L, Bruijnen RCG, et al. Staging of adenocarcinoma of the gastroesophageal junction. *Eur J Surg Oncol*. 2016;42(3):400–6.
28. Mahuron KM, Sullivan KM, Hernandez MC, et al. Diffuse-type histology is prognostic for all Siewert types of gastroesophageal adenocarcinoma. *J Gastric Cancer*. 2024;24(3):267–79.
29. Liu YT, Li YA, Li JZ, et al. Four-step formula for the Siewert classification of adenocarcinomas of the esophagogastric junction: a computed tomography-based quantitative model. *Expert Rev Med Devices*. 2025. 10.1080/17434440.2025.2510534.

Figures

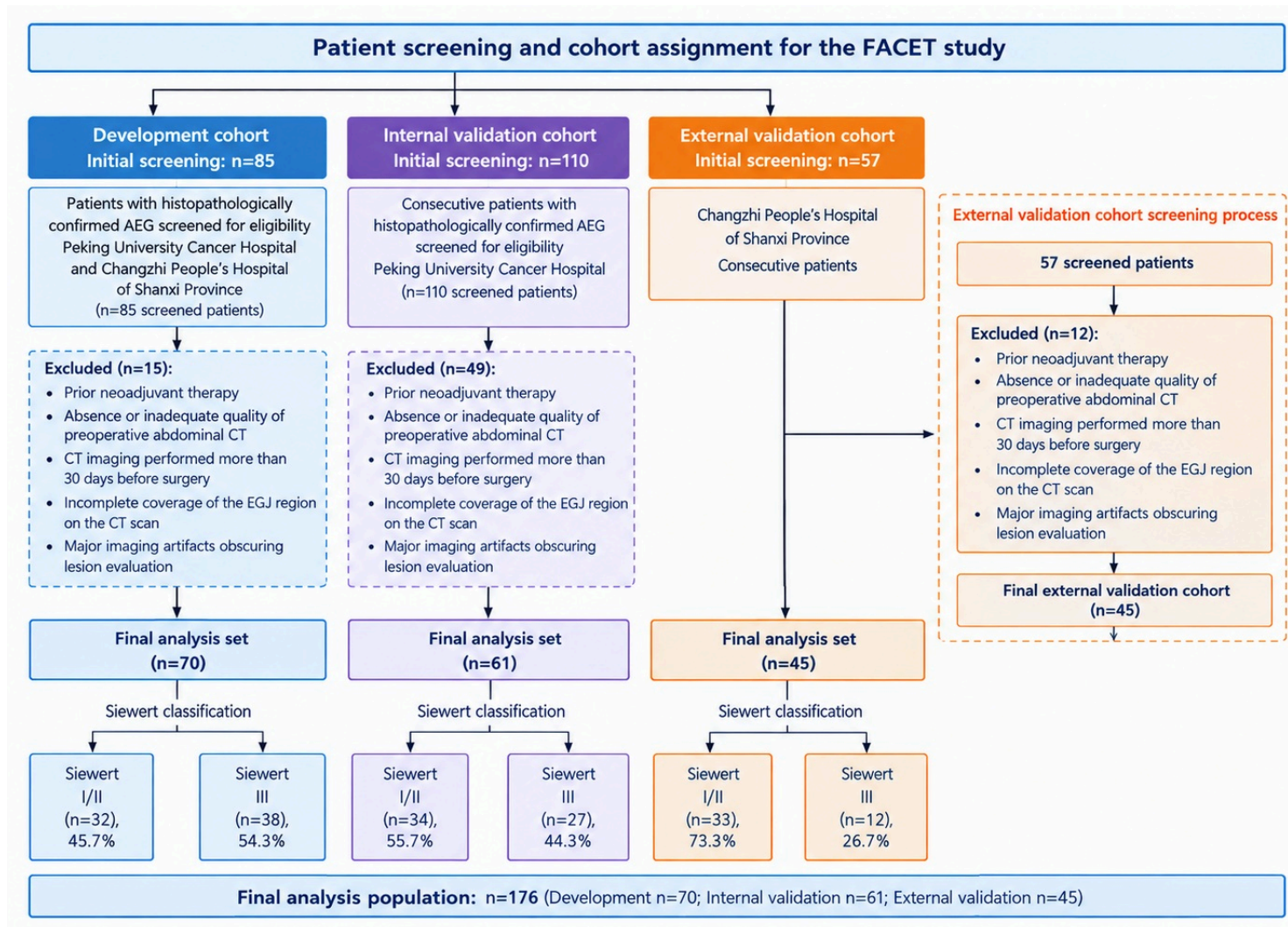


Figure 1

Patient screening, exclusions, and cohort assignment for development and validation of the FACET framework. The flow diagram summarizes the initial screening populations, application of the same prespecified exclusion criteria across all three cohorts, final analysis sets, and pathological Siewert distributions.

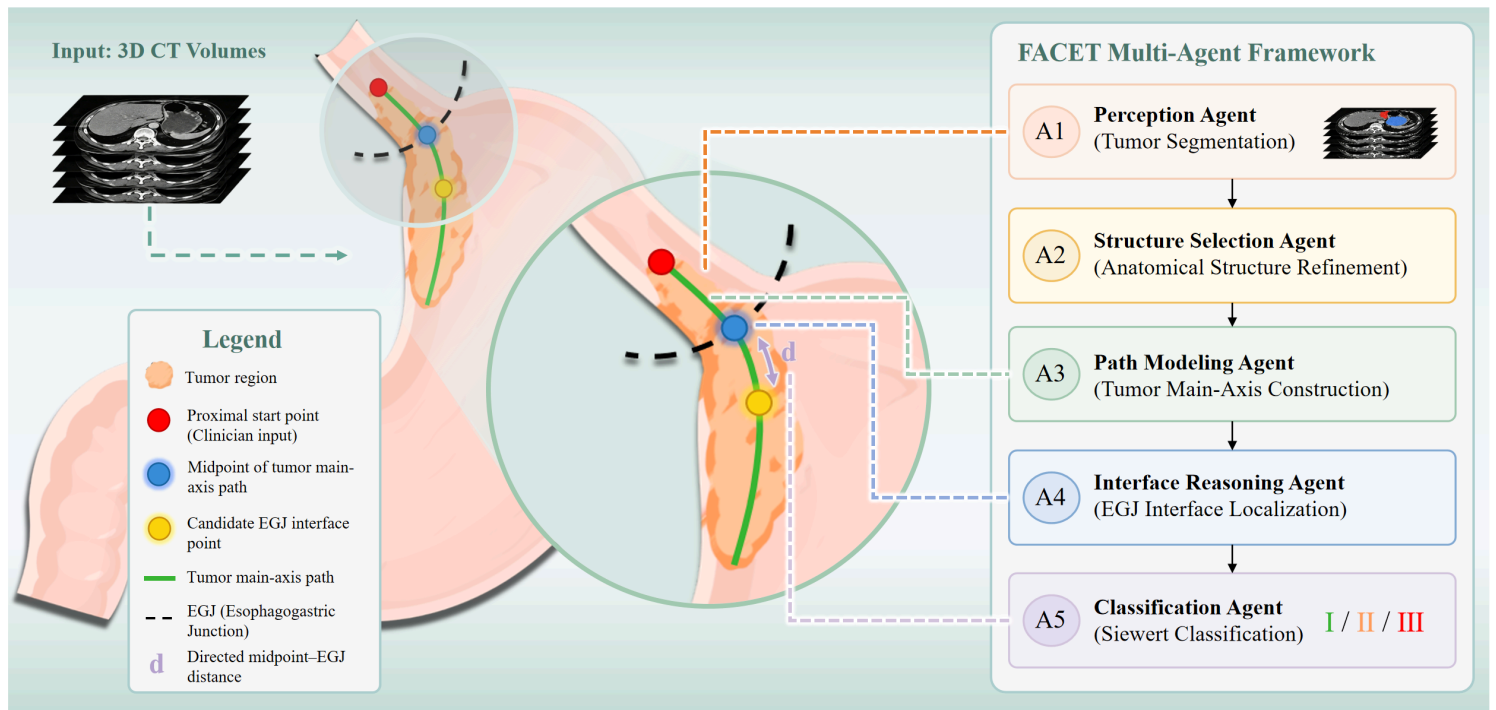


Figure 2

Overview of FACET for CT-based Siewert classification of AEG: segmentation, refinement, path modeling, EGJ localization, and distance-based typing. Legend: tumor, start, midpoint, interface point, path, EGJ, directed distance. Created using the Generic Diagramming Platform (BioGDP) [12].

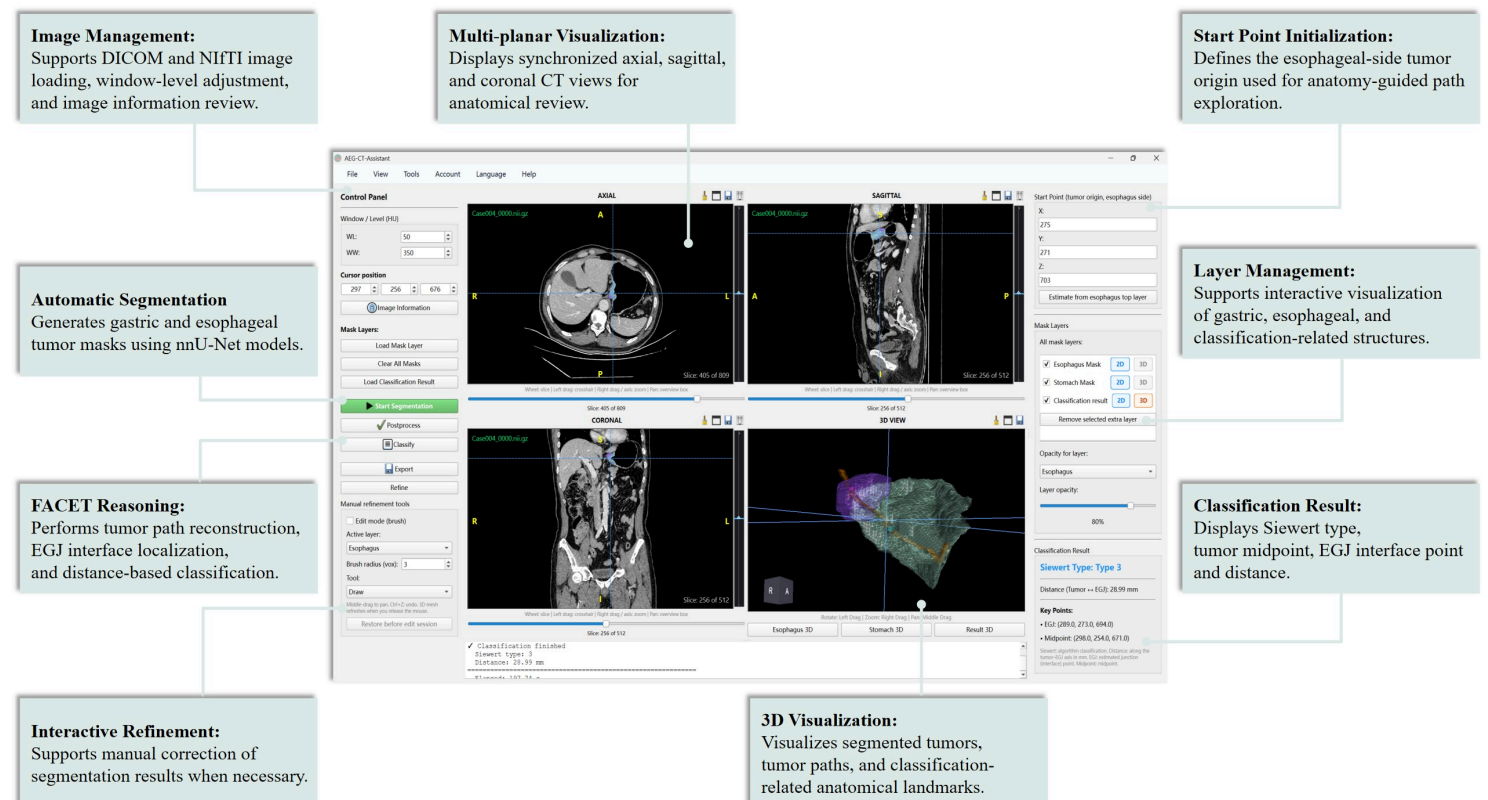


Figure 3

Clinical software implementation of the FACET framework. The platform supports image management, DICOM and NIfTI loading, window-level adjustment, multi-planar CT visualization, automated gastric and esophageal tumor segmentation, interactive refinement, start-point initialization, layer management, 3D visualization, FACET reasoning, and display of the Siewert classification result with the tumor midpoint, EGJ interface point, and distance measurement.

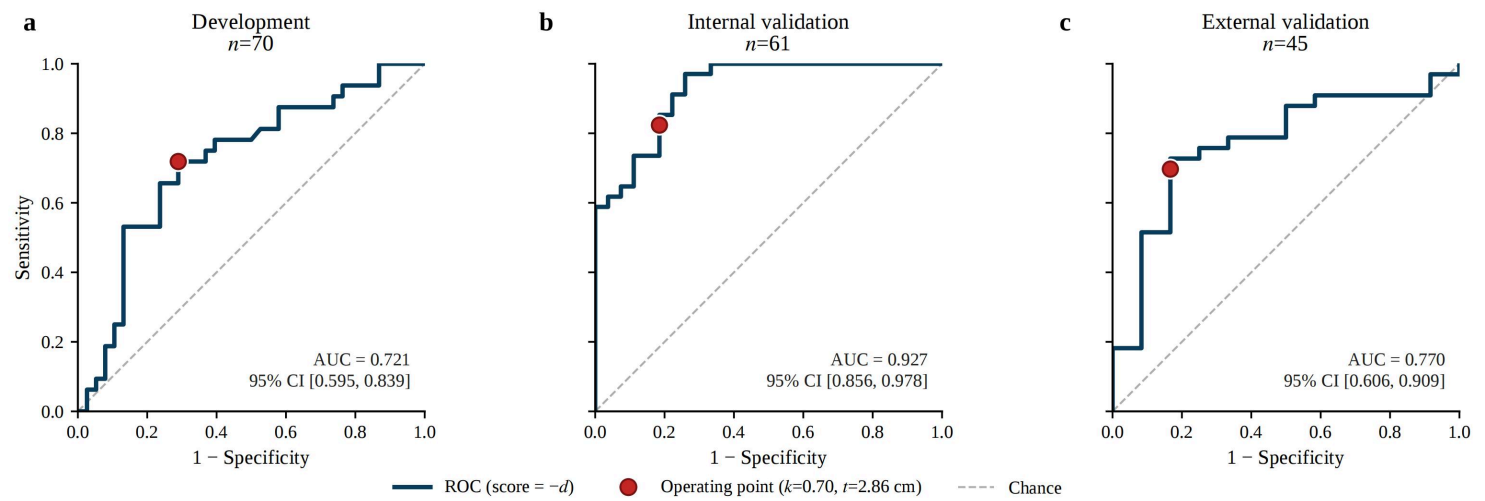


Figure 4

Receiver operating characteristic curves of FACET in the development and validation cohorts. Receiver operating characteristic curves were generated using the negative continuous calibrated distance ($-d'$) as the discrimination score, with histopathological Siewert types I/II defined as the positive class and type III as the negative class. The marked operating points correspond to the locked calibration factor $k = 0.7$ and the predefined calibrated-distance boundary of $d'=2.0$ cm, equivalent to $d=2.86$ cm on the original distance scale.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterial.docx](#)