



OPEN

Application of network analysis and association rule mining for visualizing the lymph node metastasis patterns in esophageal squamous cell carcinoma

Sangwon Baek¹, Kyunga Kim^{2,3,4,6✉}, Seong Yong Park^{5,6✉}, Yeong Jeong Jeon⁵, Junghee Lee⁵, Jong Ho Cho⁵, Hong Kwan Kim⁵, Yong Soo Choi⁵, Jae Il Zo⁵ & Young Mog Shim⁵

Understanding the patterns of lymph node (LN) metastases in esophageal squamous cell carcinoma (ESCC) is important for accurate staging and defining the extent of lymphadenectomy. This study clarified the patterns of LN metastases in ESCC using data mining techniques. 1181 patients with LN metastases who underwent upfront esophagectomy for ESCC were analyzed. Network analysis and association rule mining (ARM) techniques were employed to visualize and quantify LN metastases according to the T stage (T1 vs. T2–4) and the primary lesion location. Network plots depicted the relationship between primary lesions and metastatic LNs, and mutual LN metastasis patterns. ARM metrics assessed connection strengths among LNs. Network analysis identified the most prevalent LN metastases at 106recR/L, 105–108–110, and 1–2–7, independent of the T stage and location. ARM indicated high metastases likelihood at 106recR/L for upper ESCC, 1–2–7 and 106recR for mid-ESCC, and 1–2–7 for lower ESCC. Mutual metastases analysis identified 106recR/L, 1–2–7, and 105–108–110 as common metastasis stations across all subgroups. Conviction showed that cervical LN metastasis occurred independently of 106recR/L. Data mining techniques elucidate the intricate patterns of LN metastases and the association between metastatic LNs in ESCC.

Keywords Esophageal neoplasms, Oncologic outcome, Esophageal squamous cell cancer, Lymph node metastases, Network analysis, Association rule mining

Abbreviations

AJCC	American Joint Committee on Cancer
ARM	Association rule mining
CT	Computed tomography
ESCC	Esophageal squamous cell carcinoma
HLN	Hub lymph node
LN	Lymph node
PET	Positron emission tomography
SMC	Samsung Medical Center
101R	Right cervical paraesophageal LNs
101L	Left cervical paraesophageal LNs
102R	Right deep cervical LNs
102L	Left deep cervical LNs
104R	Right supraclavicular LNs

¹Research Institute for Future Medicine, Samsung Medical Center, Seoul, Republic of Korea. ²Biomedical Statistics Center, Research Institute for Future Medicine, Samsung Medical Center, Seoul, Republic of Korea. ³Department of Digital Health, Sungkyunkwan University, Seoul, Republic of Korea. ⁴Department of Data Convergence and Future Medicine, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea. ⁵Department of Thoracic and Cardiovascular Surgery, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea. ⁶Kyunga Kim and Seong Yong Park contributed equally to this work. ✉email: kyunga.j.kim@gmail.com; sysparkcs@gmail.com

104L	Left supraclavicular LNs
105–108–110	Thoracic paraesophageal LNs
106preR	Right pretracheal LNs
106preL	Left pretracheal LNs
106recR	Right recurrent nerve LNs
106recL	Left recurrent nerve LNs
107	Subcarinal LNs
112pulR	Right pulmonary ligament LNs
112pulL	Left pulmonary ligament LNs
1–2–7	Paracardial and left gastric artery LNs
8	Common hepatic artery LNs
9	Celiac artery LNs

Esophageal cancer is the seventh most prevalent cancer and the sixth leading cause of cancer-related mortality globally, accounting for 1 in every 18 cancer-related deaths in 2020^{1,2}. Lymph node (LN) metastases significantly influence the prognosis of patients with esophageal cancer, and the number of metastatic LNs determines the N factors in the current 8th edition of the American Joint Committee on Cancer (AJCC) staging system³. Accurate assessment of metastatic LNs in esophageal cancer requires a thorough understanding of the patterns of LN metastases. Owing to the unique anatomical characteristics of the esophagus, such as its distinct submucosal lymphatic drainage system with longitudinal networks, esophageal cancer can metastasize from the neck to the abdomen even in the early stages^{4,5}. Furthermore, the distribution of LN metastases may differ based on clinicopathological factors, including the location of the primary lesion and invasion depth (T). Previous studies have focused on examining metastatic patterns by reporting the frequency of metastases in specific regions^{6,7}. Although these analyses have provided a general overview of metastatic prevalence, they have failed to delineate the intricate and detailed distribution patterns of LN metastases to specific LN stations⁷.

Network analysis is a widely used technique for examining the relationships between discrete objects^{8,9}. Traditionally employed in communication or social network analysis, this theory has recently found applications in oncology^{8,9}. In addition, association rule mining (ARM), frequently used in market basket analysis, is a robust methodology for identifying the strength of connections among discrete objects within large datasets¹⁰. We hypothesized that employing network analysis and ARM for esophageal squamous cell carcinoma (ESCC) could offer a clear visualization and quantifiable interpretation of the complex LN metastasis patterns in ESCC. Consequently, this study aimed to investigate and visualize the precise patterns of LN metastases in patients with ESCC, which is prevalent in East Asia¹¹, using network analysis and ARM.

Methods

Ethical considerations

This research was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study protocol was approved, and the requirement for informed consent was waived by the institutional review board of Samsung Medical Center (No. SMC 2023-10-091).

Patients

This retrospective observational study was conducted at the Samsung Medical Center (SMC), South Korea. The study cohort was extracted from the Registry for Thoracic Cancer Surgery at SMC, which included 2663 patients who underwent esophagectomy for ESCC between 1994 and 2022. Patients who either received preoperative therapy or had pathologies other than ESCC were excluded. Finally, 1181 patients with recognizable LN metastases were included in the network construction and ARM.

Esophagectomy and intrathoracic or cervical esophagogastrostomy were performed in most patients. Three-field LN dissection is routinely performed for patients with upper thoracic esophageal cancer. When preoperative computed tomography (CT) or positron emission tomography (PET)-CT suggested cervical LN metastases in patients with mid- to lower thoracic esophageal cancer, histological confirmation by ultrasound-guided needle aspiration was conducted whenever possible. However, if cervical LN metastases are not detected on both CT and PET-CT, neck dissection is not performed in patients with mid- to lower thoracic esophageal cancer. Regarding mediastinal LN dissection, routine upper mediastinal LN dissection, including bilateral recurrent laryngeal nerve nodes (106recR and 106recL), was performed¹².

The LN stations were labeled using the station numbers specified in the 12th Japanese Guidelines for Clinical Pathological Studies on Carcinoma of the Esophagus¹³. In our institutional procedures, esophagectomy and lymphadenectomy were performed en bloc; paracardial (1 and 2) and left gastric (7) LNs were dissected together along the stomach. In addition, all paraesophageal LNs were dissected with the esophagus as attached, although the Japanese classification divided it into upper (105), mid- (108), and lower paraesophageal LNs (110). Therefore, the name of the LN station was modified, and the terminologies “1–2–7” and “105–108–110” were used in the analysis.

Visualization of the lymphatic metastasis network

Network analyses were conducted to elucidate the complex patterns of LN metastases in patients with ESCC. To examine the clinical implications of tumor depth and the location of the primary lesion, patients were stratified into six subgroups according to the T stage (T1 vs T2–4) and location of the primary lesion (upper, mid, and lower). For each subgroup, three networks were developed according to the N stage. Under the assumption that cancer cells drain to regional LNs in similar pathways, three networks were explored to observe the metastatic patterns per the different extent of metastases. The first network was constructed based on metastatic LNs

dissected from patients with N1 (LNs ≤ 2). The second and third networks comprised metastatic LNs dissected from patients with N1–N2 (LNs ≤ 6) and N1–N3 (all LNs), respectively. Unique findings from these networks were displayed in a single network map by color-coding the metastatic LNs (yellow for N1, light green for N1–N2, and dark green for N1–N3), which are presented as circular vertices in the plot.

The purpose of the network analysis was two-fold. First, network analysis was applied to display all metastatic LNs arising from the primary tumor. The size of each vertex was set based on the number of patients having corresponding metastatic LNs from the primary lesion, with larger vertices implying more commonly metastatic LNs. Then, these vertices were color-coded according to the N stage at which metastatic LNs firstly occurred to contrast the extent of tumor cell dissemination to specific LN stations. Such visualization may provide insights into sequential information regarding the lymphatic flow of tumor cells from the primary lesion. The networks of the six subgroups were superimposed on the esophagus maps to demonstrate the locations of metastatic LNs. This additional information on anatomical locations to color-coding may help clinicians explore the course of metastatic tumor progression.

Second, network analysis of patients with multiple metastatic LNs was performed to assess the association between LNs and identify influential LNs in the network. LNs dissected from the same patient were connected with lines, which are referred to as edges. For each LN pair, a nodal connection was identified if the LNs were dissected from the same patient. The tie strength for an LN pair was defined as the ratio of nodal connections between the LN pair to the total nodal connections between all LN pairs within the network¹⁴. The thickness of each edge was determined by its tie strength, with thicker edges indicating stronger nodal interconnectedness. As a measure of vertex centrality, the degree of an LN was defined as the number of connections identified for the LN¹⁵.

By integrating the degree with the tie strength, the weighted degree was further calculated to provide a comprehensive measure of the significance of an LN within the network¹⁵. Then, the weighted degree for each LN was derived by aggregating the tie strengths of all connections related to that LN. The vertex size for each LN represents its weighted degree. In addition, hub LNs (HLNs) were defined as LNs with a weighted degree ≥ 0.2 , demonstrating broader and stronger linkages with other LNs. These HLN s were distinguished by assigning colors based on the N stage at which they were selected, whereas all remaining LNs were colored gray.

Assessment of metastatic pattern strength using ARM

ARM, a data mining technique for uncovering significant relationships across large datasets, was performed to quantify the important associations among metastasized LNs and assess the associative strengths for discovering metastatic patterns using data from 1181 patients¹⁶. In ARM, a transaction is defined as a set of metastatic LNs dissected from a patient. An item represents either a location of the primary lesion or an LN station, and an item set is a collection of one or more items. An association rule is identified if the two item sets occur together in the transaction. These item sets are classified either as the antecedent ("if" part) or as the consequent ("then" part). We identified all possible association rules and evaluated their significance using three ARM metrics: support, confidence, and conviction. The definition and formula of each metric are shown as follows.

- (1) Support measures the frequency of two items co-occurring in transactions.

$$\text{Support } (X, Y) = \frac{\text{Frequency}(X, Y)}{N}$$

- (2) Confidence is the likelihood of selecting a consequent item when its antecedent is chosen.

$$\text{Confidence } (X \rightarrow Y) = \frac{\text{Frequency}(X, Y)}{\text{Frequency}(X)}$$

- (3) Conviction evaluates the dependency of a consequent item's occurrence on the presence of an antecedent item.

$$\text{Conviction } (X \rightarrow Y) = \frac{1 - \text{Support } (Y)}{1 - \text{Confidence } (X \rightarrow Y)}$$

Visualization of the ARM results

The results of the three ARM metrics were visualized using heatmaps and tree graphs. First, the prevalence of metastatic LNs (consequent) from the primary tumor (antecedent) was assessed using the confidence metric. According to the T stage and location of the primary lesion, the confidence values from the primary sites to their metastatic LNs with high support were displayed based on a predefined color scheme in the heatmaps. Dark and light red were applied to signify strong and weak confidence levels, respectively. To highlight the most relevant associative rules, only confidence values > 0.05 were presented.

Second, all LNs codissected with HLN s identified from previous network analysis were discovered using the support metric. Similarly, the support values were displayed based on a predefined color scheme in heatmaps according to the T stage and anatomical location of the primary tumor. Only support values ≥ 0.02 were given to represent the significant nodal connections. In addition, a bold vertical line was drawn to distinguish between HLN s and non-HLN s.

Third, conviction metric was used to assess the likelihood of dependent occurrence LNs connected to 106recR and 106recL. Conviction values from 106recR and 106recL (antecedents) to LN stations (consequents) were displayed in tree graphs. The vertex size indicated the support value for each LN, which measures the frequency of each LN occurring in all transactions. The thickness of the edge indicates the tie strength between two LNs. Conviction values ranged from 0.84 to 1.04. A conviction value close to 1 implies that metastases to consequent LN stations occur independently of metastases to antecedent LN stations.

Statistical analysis

Patient characteristics were summarized as mean \pm standard deviation and number (%) for continuous and categorical variables. Their group comparisons were conducted using the independent t-test and Chi-squared contingency test for continuous and categorical variables, respectively. A p -value < 0.05 was considered statistically significant. All analyses and visualizations were performed using Python (Python Software Foundation, version 3.9).

Results

Patient characteristics

Of the 1,181 patients with ESCC, 391 (33.1%) had pT1 tumors and 790 (66.9%) had pT2–4 tumors (Table 1). The locations of the primary lesions were similarly distributed between the pT1 and pT2–4 groups. A higher proportion of patients with pT1 tumors were classified as pN1 (75.4% vs. 45.6%), whereas those with pT2–4 tumors had a higher proportion of pN2 (35.9% vs. 21.5%) and pN3 (18.5% vs. 3.1%) than patients with pT1 tumors ($p < 0.001$). More LNs were dissected in the pT2–4 group than in the pT1 group (44.71 ± 19.41 vs. 37.67 ± 14.44 , $p < 0.001$).

The prevalence of metastases across the LN stations is detailed in Table 2. For upper ESCC, the highest prevalence was observed in 106recR (51.06% in pT1 and 54.88% in pT2–4), followed by 106recL (27.66% in pT1 and 48.78% in pT2–4). In mid-ESCC, 1–2–7 exhibited the most frequent prevalence (33.68% in pT1 and 48.13% in pT2–4), followed by 106recR (34.2% in pT1 and 37.17% in pT2–4). In lower ESCC, 1–2–7 showed the highest prevalence (62.91% in pT1 and 66.47% in pT2–4), followed by 105–108–110 (15.89% in pT1 and 37.72% in pT2–4).

Patterns of lymphatic metastases from primary lesion sites to specific LN stations

The LN metastatic patterns according to the invasion depth (T1 vs T2–4) and location of the primary lesions were visualized through the network plot (Fig. 1). The sizes of each node signified the magnitude of their connection to their respective primary sites. The highest prevalence of LN metastases was at 106recR, 106recL, 105–108–110, and 1–2–7, regardless of the invasion depth and location. In patients with upper esophageal cancer, LN metastases were also observed at 104R, 104L, 102R, and 102L. In patients with mid-ESCC, LN metastases were also detected at 104R, 104L, 102R, 102L, 106preR, and 112pull/R. T1 ESCC was highly prevalent in 106recR, 106recL, and 1–2–7, whereas T2–4 ESCC was highly prevalent in 105–108–110 and 107 regardless of location. Metastases to the abdominal LNs such as 9 in upper ESCC and cervical LNs such as 104R and 104L in lower ESCC were observed in advanced N stages (N2 or N3).

	pT1 (n = 391)	pT2–4 (n = 790)	p
Age	62.50 ± 8.15	64.82 ± 8.40	$< .001$
Male	370 (94.6%)	754 (95.4%)	0.638
Total numbers of dissected LNs	37.67 ± 14.44	44.71 ± 19.41	$< .001$
Location of the primary lesion			0.417
Upper	47 (12.0%)	82 (10.4%)	
Mid	193 (49.4%)	373 (47.3%)	
Lower	151 (38.6%)	334 (42.3%)	
T stage			$< .001$
pTis and pT1a	37 (9.5%)	0 (0.0%)	
pT1b	354 (90.5%)	0 (0.0%)	
pT2	0 (0.0%)	195 (24.7%)	
pT3	0 (0.0%)	538 (68.1%)	
pT4a	0 (0.0%)	22 (2.8%)	
pT4b	0 (0.0%)	35 (4.4%)	
N stage			$< .001$
pN1	295 (75.4%)	360 (45.6%)	
pN2	84 (21.5%)	284 (35.9%)	
pN3	12 (3.1%)	146 (18.5%)	

Table 1. Patient demographics and characteristics. Data are presented as mean \pm standard deviation or N (proportion in %) LN lymph nodes.

	Upper		Mid		Lower	
	pT1	pT2–4	pT1	pT2–4	pT1	pT2–4
	(n=47)	(n=82)	(n=193)	(n=374)	(n=151)	(n=334)
Cervical region						
101R (right cervical paraesophageal LNs)	3 (6.38%)	7 (8.54%)	4 (2.07%)	11 (2.94%)	1 (0.66%)	1 (0.3%)
101L (left cervical paraesophageal LNs)	1 (2.13%)	8 (9.76%)	1 (0.52%)	15 (4.01%)	0 (0.0%)	2 (0.6%)
102R (right deep cervical LNs)	0 (0.0%)	2 (2.44%)	0 (0.0%)	1 (0.27%)	0 (0.0%)	0 (0.0%)
102L (left deep cervical LNs)	0 (0.0%)	1 (1.22%)	1 (0.52%)	1 (0.27%)	0 (0.0%)	0 (0.0%)
104R (right supraclavicular LNs)	1 (2.13%)	9 (10.98%)	1 (0.52%)	10 (2.67%)	0 (0.0%)	2 (0.6%)
104L (left supraclavicular LNs)	4 (8.51%)	7 (8.54%)	0 (0.0%)	5 (1.34%)	0 (0.0%)	2 (0.6%)
Mediastinal region						
106preR (right pretracheal LNs)	0 (0.0%)	0 (0.0%)	1 (0.52%)	7 (1.87%)	0 (0.0%)	2 (0.6%)
106preL (left pretracheal LNs)	0 (0.0%)	0 (0.0%)	1 (0.52%)	12 (3.21%)	0 (0.0%)	5 (1.5%)
106recR (right recurrent nerve LNs)	24 (51.06%)	45 (54.88%)	66 (34.2%)	139 (37.17%)	27 (17.88%)	64 (19.16%)
106recL (left recurrent nerve LNs)	13 (27.66%)	40 (48.78%)	49 (25.39%)	109 (29.14%)	22 (14.57%)	42 (12.57%)
107 (subcarinal LNs)	2 (4.26%)	8 (9.76%)	8 (4.15%)	93 (24.87%)	8 (5.3%)	56 (16.77%)
105–108–110 (thoracic paraesophageal LNs)	6 (12.77%)	20 (24.39%)	32 (16.58%)	140 (37.43%)	24 (15.89%)	126 (37.72%)
112pulR (right pulmonary ligament LNs)	0 (0.0%)	0 (0.0%)	1 (0.52%)	12 (3.21%)	3 (1.99%)	6 (1.8%)
112pulL (left pulmonary ligament LNs)	0 (0.0%)	1 (1.22%)	4 (2.07%)	20 (5.35%)	11 (7.28%)	29 (8.68%)
Abdominal region						
1–2–7 (paracardial and left gastric artery LNs)	4 (8.51%)	12 (14.63%)	65 (33.68%)	180 (48.13%)	95 (62.91%)	222 (66.47%)
8 (common hepatic artery LNs)	1 (2.13%)	1 (1.22%)	2 (1.04%)	19 (5.08%)	2 (1.32%)	32 (9.58%)
9 (celiac artery LNs)	1 (2.13%)	1 (1.22%)	7 (3.63%)	31 (8.29%)	14 (9.27%)	53 (15.87%)

Table 2. Number of dissected lymph nodes at each station according to the T stage and location of the primary lesion. LN lymph node, n represents the number of patients.

The connection strength from each primary tumor to the LN stations was measured using confidence values and displayed through a two-dimensional heatmap (Fig. 2). In this plot, the prevalence of LN metastases was depicted from the primary sites (antecedents) to each LN station (consequents) according to the number of metastatic LNs using confidence values that indicate the strength of each association. In patients with upper ESCC, 106recR and 106recL were the prevalent stations with high confidence values (≥ 0.2) from one metastatic LN, and the confidence value of 106recR was slightly decreased, which might be related to the increased number of metastatic LNs. In addition, the confidence values gradually increased at 105–108–110 in patients with pT1 (ranged 0.04–0.13) and pT2–4 (ranged 0.11–0.26) tumors as the number of metastatic LNs increased. In patients with mid-ESCC, 1–2–7 and 106recR were the most frequent stations with high confidence values (≥ 0.2), followed by 105–108–110 (pT1, 0.10–0.17; pT2–4, 0.22–0.38). Moreover, patients with pT2–4 mid-ESCC had a higher confidence (0.07–0.25) to have metastases at 107 than patients with pT1 (0.01–0.04). Patients with lower ESCC were more likely to have metastases at 1–2–7 with substantial confidence (pT1, 0.51–0.63; pT2–4, 0.53–0.67), followed by 105–108–110 LNs (pT1, 0.13–0.16; pT2–4, 0.20–0.38). Interestingly, if the numbers of metastatic LNs were small (N1), the confidences values were high in 106recR, 106recL, and 1–2–7, and as the numbers of metastatic LNs increased, the confidence values of other LN stations increased.

Patterns of mutual metastases among LN stations

Nodal connections were further analyzed using the network plot to discover patterns of mutual metastases among LN stations (Fig. 3). Despite some variations, 106recR, 106recL, 1–2–7, and 105–108–110 were commonly selected as HLNs with the highest degree across all subgroups. In particular, in patients with upper ESCC, 104L was selected as the HLN instead of 1–2–7. In patients with mid-ESCC, 107 were also selected as the HLN. In patients with lower ESCC, those with pT1 ESCC had 9 as the HLN and those with pT2–4 ESCC had 107 as the HLN. Mutual metastases with high tie strength were more prevalent in patients with pT2–4 ESCC than in those with pT1 ESCC.

Patterns of mutual metastases were quantitatively assessed using support values for LN pairs and plotted in a two-dimensional heatmap (Fig. 4). Irrespective of the primary lesion, patients with pT1 ESCC generally had mutually limited metastases within HLNs. In contrast, some HLNs in pT2–4 ESCC were likely to have mutual metastases with other non-HLNs. In patients with a pT2–4 in the upper ESCC, 106recR/L tended to have mutual metastases with 101R, 101L, 107, 1–2–7 (support ≥ 0.03). In patients with pT2–4 mid-ESCC, most HLNs showed mutual metastases with 9 (support ≥ 0.03). In patients with pT2–4 lower ESCC, 1–2–7 also metastasized with 106recL, 112pulR, and 8.

The mutual metastatic patterns between 106recR/L and other frequently co-occurring LNs, including cervical LNs (101R/L, 104R/L), were further investigated to see whether metastases of those LNs are likely to follow the metastases of 106recR/L (Fig. 5). In upper ESCC, all conviction values from 106recR/L to other LNs were very close to 1 (i.e., ranged 0.90–1.04). It can imply that metastases to other LNs occurred independently of 106recR/L. This was observed also in mid ESCC (conviction values: 0.84–1.03).

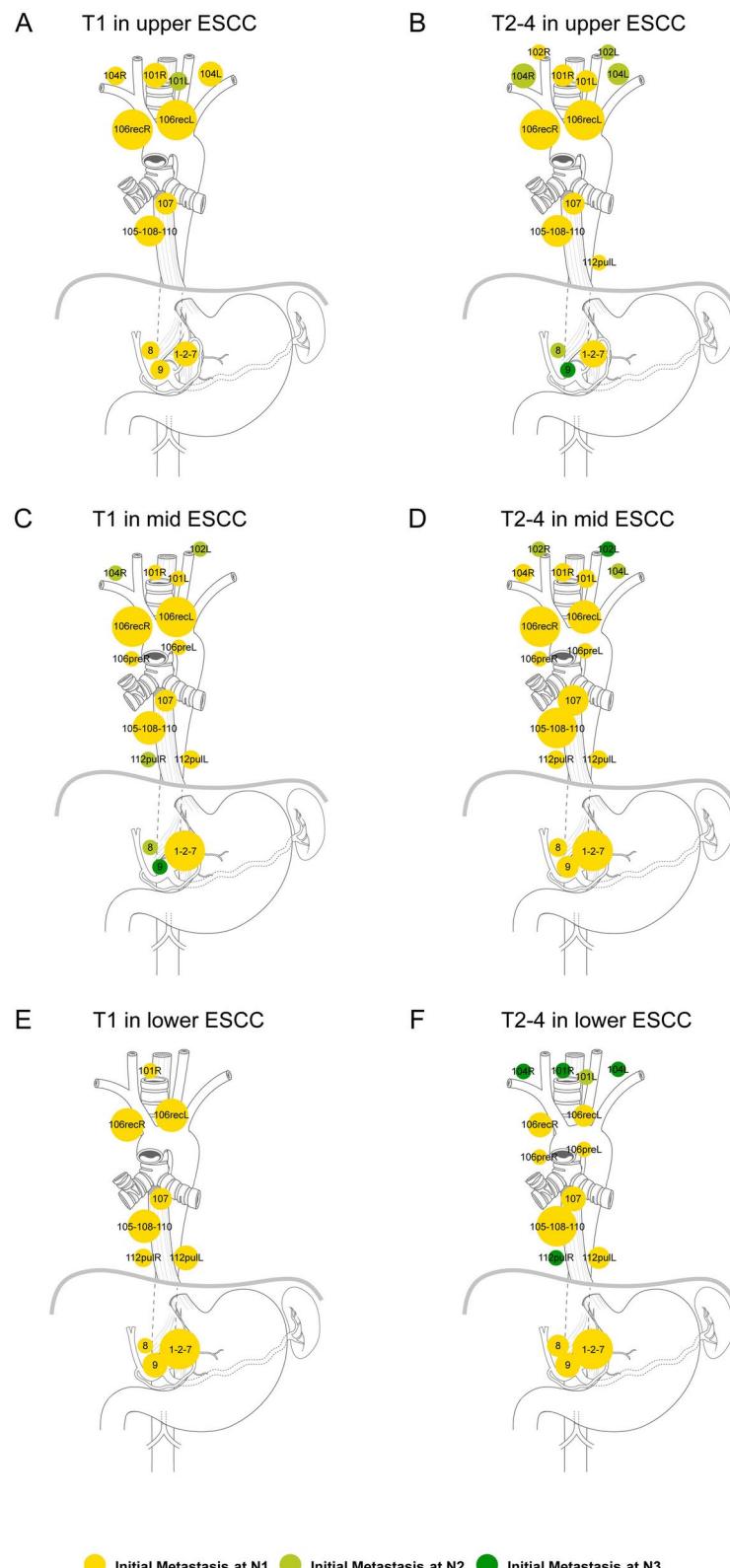


Fig. 1. Network of lymph node (LN) metastases from the primary lesion to specific LN stations. **(A)** T1 ESCC in the upper esophagus. **(B)** T2–4 ESCC in the upper esophagus. **(C)** T1 ESCC in the mid-esophagus. **(D)** T2–4 ESCC in the mid-esophagus. **(E)** T1 ESCC in the lower esophagus. **(F)** T2–4 ESCC in the lower esophagus. *The size of each vertex reflects the frequency of lymphatic metastases at the corresponding LN station. The vertices were color-coded according to the N stage to represent the LNs that also metastasized at advanced N stages.

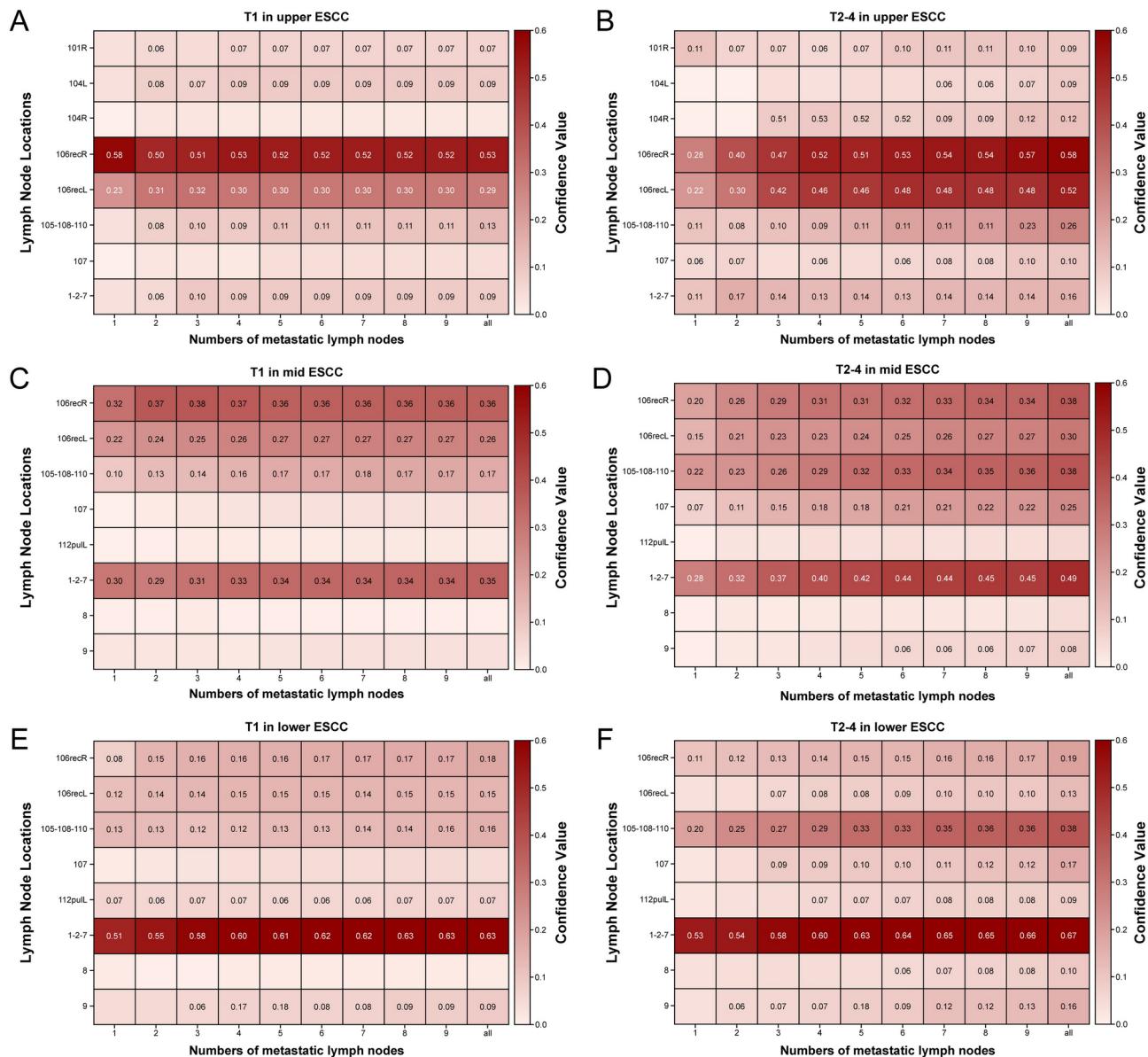


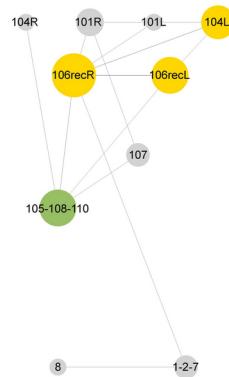
Fig. 2. Trend in lymph node (LN) metastases based on the number of metastatic LNs. (A) T1 ESCC in the upper esophagus. (B) T2-4 ESCC in the upper esophagus. (C) T1 ESCC in the mid-esophagus. (D) T2-4 ESCC in the mid-esophagus. (E) T1 ESCC in the lower esophagus. (F) T2-4 ESCC in the lower esophagus. *Frequently metastasized LNs (top 8) were selected and displayed on the y-axis. Subgroup analysis results for LNs classified according to the number of metastatic LNs are displayed on the x-axis. Degree of confidence values are expressed through a red color scheme. Dark red indicates high confidence, and light red indicates low confidence.

Discussion

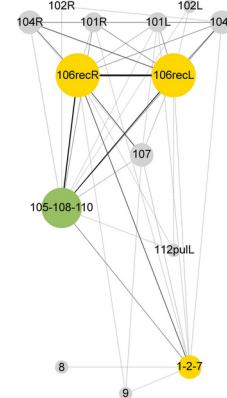
In this study, network analysis and ARM were employed to visualize LN metastases in ESCC, according to the invasion depth and primary lesion location, and quantify the associations between LN stations. The results were similar to those of previous studies, which represented patterns of LN metastases in ESCC by simply showing the frequency of LN metastases^{6,17,18}, however, we have built upon them by applying network analysis and ARM with mathematically based quantitative metrics. This approach has allowed us to not only visualize but also quantify the more intricate patterns of LN metastases in ESCC.

Understanding the patterns of LN metastases is crucial for the management of all cancers, including esophageal cancer. In particular, in ESCC, recognizing metastasis patterns may allow for the determination of an appropriate extent of lymphadenectomy, which may enable accurate staging and potential therapeutic benefits¹⁹. The unique lymphatic drainage system of the esophagus, which includes longitudinal networks within the submucosal layers, allows for the metastases of cancer cells from the neck to the abdomen via the submucosal lymphatic network^{4,5}. In addition, previous studies have reported that LN metastases frequently occur in

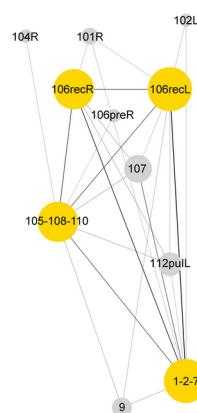
A T1 in upper ESCC



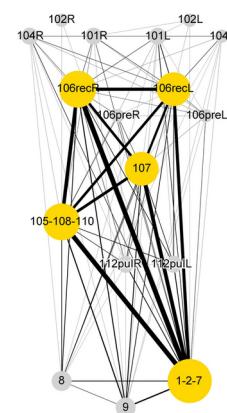
B T2-4 in upper ESCC



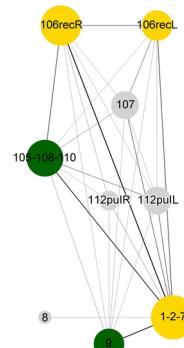
C T1 in mid ESCC



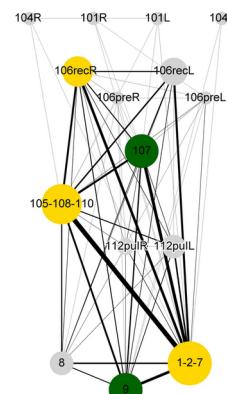
D T2-4 in mid ESCC



E T1 in lower ESCC



F T2-4 in lower ESCC



● Initial Metastasis at N1 ● Initial Metastasis at N2 ● Initial Metastasis at N3

Fig. 3. Visualization of network analysis of lymph node (LN) metastasis between LN stations. (A) T1 upper ESCC. (B) T2-4 ESCC in the upper esophagus. (C) T1 ESCC in the mid-esophagus. (D) T2-4 ESCC in the mid-esophagus. (E) T1 ESCC in the lower esophagus. (F) T2-4 ESCC in the lower esophagus. *The sizes of each vertex reflect its weighted degree within the network. The edge thickness reflects the tie strength between the two LNs. Hub LNs were color-coded according to the N stage at which they were selected.

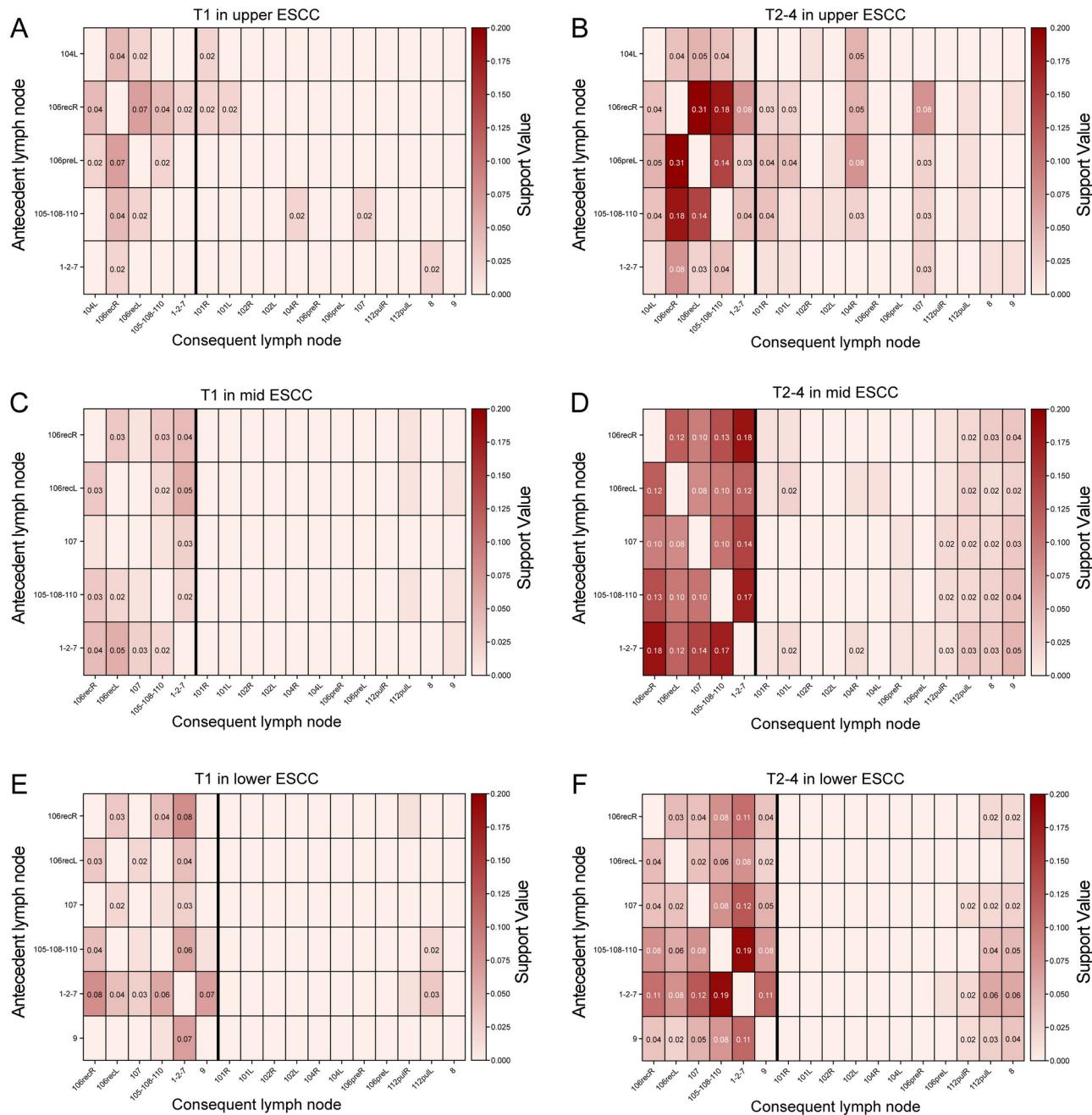


Fig. 4. Patterns of mutual lymph node (LN) metastases between hub and other LNs. (A) T1 ESCC in the upper esophagus. (B) T2-4 ESCC in the upper esophagus. (C) T1 ESCC in the mid-esophagus. (D) T2-4 ESCC in the mid-esophagus. (E) T1 ESCC in the lower esophagus. (F) T2-4 ESCC in the lower esophagus. *A heatmap was drawn with the LNs dissected from the entire patient cohort (N1-N3). Hub LNs were identified and displayed on the y-axis. All LNs codissected with hub LNs were discovered and are displayed on the x-axis. The degree of support values are expressed through a red color scheme. Dark red indicates high support, and light red indicates low support. The vertical line was drawn to segment the hub and non-hub LNs.

106recR, 106recL, and cervical LNs in ESCC²⁰⁻²². As a result, 106recR and 106recL dissections and three-field LN dissection have been suggested as essential surgical procedures. However, such extensive LN dissections increase the risk of postoperative complications, such as vocal cord palsy and pulmonary complications²³; consequently, controversies remain between three- and two-field LN dissections regarding postoperative outcomes and survival benefits²⁴.

This analysis was initially designed to explore the sequences and patterns of LN metastases in resected ESCC using mathematical and statistical models. However, the pathological reports after surgery provided only the results of LN metastases, omitting the sequence of LN involvement. Consequently, the construction of a mathematical sequence model was unfeasible. Instead, we developed a model that determines the confidence

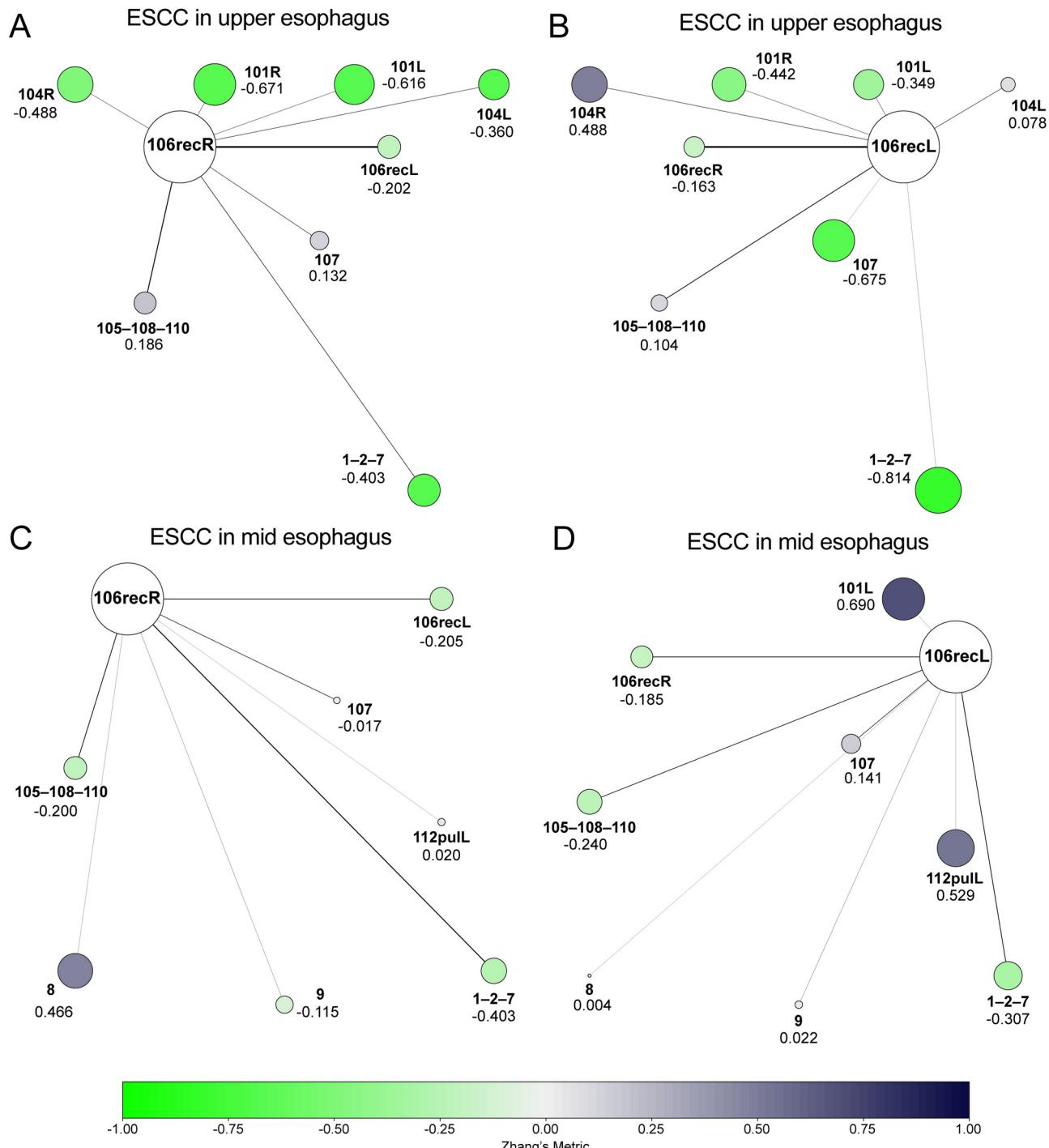


Fig. 5. Visualization of network analysis using conviction. **(A)** Relationship between 106recR and other LNs in upper ESCC. **(B)** Relationship between 106recL and other LNs in upper ESCC. **(C)** Relationship between 106recR and other LNs in mid-ESCC. **(D)** Relationship between 106recL and other LNs in mid-ESCC. *Vertex sizes were determined based on each individual LN's support values with respect to all transactions. Edge width indicates the tie strength between paired LNs.

value for each LN station by considering the number of metastatic LNs (Fig. 2). According to this model, metastatic LNs in patients with fewer metastases (early N stage) are statistically considered the initial sites of metastases. The LN stations of 106recR, 106recL, and 1–2–7 exhibited high confidence values at an early N stage, independent of the T stage and location. These results suggest that early metastases are likely to occur at 106recR, 106recL, and 1–2–7 stations, with metastases to other LN stations typically occurring at a more advanced N stage. This inference aligns with the anatomical study by Mizutani et al., which demonstrated a morphological

link between the submucosal lymphatic vessels of the proximal esophagus and 106recR/L; thus, 106recR/L are frequently the initial sites of LN metastases because of these anatomical characteristics²⁵.

Another intriguing aspect of this study was the observed differences in metastatic LN patterns between T1 and T2–4 lesions. In patients with upper ESCC, the confidence values in LN stations 106recR and 106recL consistently increased for both pT1 and pT2–4 tumors. Patients with mid-ESCC exhibited similar rising trends in confidence values for the 106recR, 106recL, and 1–2–7 stations. In contrast, patients with lower ESCC demonstrated increased confidence values specifically in the 1–2–7 station. In addition, in patients with pT2–4 tumors across all primary locations, the confidence values in 105–108–110 stations also increased, with these levels being higher than those associated with T1 lesions. These observations echo the findings of Tachimori et al., who noted that LN metastases were frequently observed in the upper mediastinum and perigastric area in patients with submucosal tumors in the mid- and lower esophagus, facilitated by the submucosal lymphatic network⁶. They also reported an increased frequency of periesophageal LN metastases in the mid- and lower mediastinum when the tumor invaded or penetrated the muscle layer, indicating a more advanced ESCC⁶. A key distinction of our study from previous studies is our calculation of the confidence values at individual LN stations rather than merely reporting the frequency of metastases in broader regions such as the upper, mid-, and lower mediastinum.

The exploration of concurrent metastatic patterns among LN station pairs suggested that 106recR and 106recL frequently metastasized with cervical LNs, such as 101R, 101L, 104R, and 104L, especially in patients with pT2–4 upper ESCC. Considering that three-field LN dissection is associated with increased postoperative complications, some studies have proposed using the absence of metastases at 106recR and 106recL as an indicator to potentially forgo neck dissection, assuming that it could predict the lack of cervical LN metastases^{26,27}. However, our analysis with ARM did not reveal that the metastases in cervical LN stations were likely to follow those in 106recR and 106recL. In other words, there was no evidence that the metastases of 106recR and 106recL can be precursory to those of the cervical LNs.

This study has several limitations. First, although data from over 1000 patients were analyzed, this number is still insufficient for comprehensive network analysis. Second, as noted in the Methods section, esophagectomy and lymphadenectomy were performed en bloc by following our institution's policies; this meant that paracardial and left gastric LNs were dissected together. In addition, paraesophageal LNs were removed along with the esophagus; consequently, 105–108–110 stations were not individually separated. Third, three-field LN dissection was not performed in all patients, particularly those with mid- to lower esophageal cancer without clinical cervical LN metastases. Therefore, the confidence and support values for cervical LNs in patients with mid- to lower ESCC could be underestimated. Despite these limitations, to the best of our knowledge, this is the first study to apply network analysis and ARM to ESCC research. The significance of this study lies in the introduction of a novel methodology for the analysis of LN metastases in ESCC. The application of these methods to larger datasets, such as the Japanese registry data, could yield further insights into LN metastases in ESCC.

This retrospective study employed data mining techniques to delineate the patterns of LN metastases in ESCC. These patterns varied according to the invasion depth and primary lesion location. Conviction indicated that 106recR/L are not suitable surrogate markers for cervical metastases. Future research using larger datasets with detailed documentation of LN stations is anticipated to uncover new perspectives on LN metastases in ESCC.

Data availability

This study used deidentified patient data extracted from the Registry for Thoracic Cancer Surgery at SMC. The data sets analyzed during this study are available from the corresponding author on reasonable request.

Received: 23 June 2024; Accepted: 4 February 2025

Published online: 13 February 2025

References

1. Sung, H. et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**(3), 209–249 (2021).
2. Park, S. Y. & Kim, D. J. Esophageal cancer in Korea: Epidemiology and treatment patterns. *J. Chest Surg.* **54**(6), 454–459 (2021).
3. Rice, T. W., Patil, D. T. & Blackstone, E. H. 8th edition AJCC/UICC staging of cancers of the esophagus and esophagogastric junction: Application to clinical practice. *Ann. Cardiothorac. Surg.* **6**(2), 119–130 (2017).
4. Riquet, M. et al. Lymphatic drainage of the esophagus in the adult. *Surg. Radiol. Anat.* **15**(3), 209–211 (1993).
5. Kuge, K. et al. Submucosal territory of the direct lymphatic drainage system to the thoracic duct in the human esophagus. *J. Thorac. Cardiovasc. Surg.* **125**(6), 1343–1349 (2003).
6. Tachimori, Y. et al. Pattern of lymph node metastases of esophageal squamous cell carcinoma based on the anatomical lymphatic drainage system. *Dis. Esophagus* **24**(1), 33–38 (2011).
7. Hagens, E. R. C., van Berge Henegouwen, M. I. & Gisbertz, S. S. distribution of lymph node metastases in esophageal carcinoma patients undergoing upfront surgery: A systematic review. *Cancers (Basel)* **12**(6), 1592 (2020).
8. Chen, L. L. et al. Cancer metastasis networks and the prediction of progression patterns. *Br. J. Cancer* **101**(5), 749–758 (2009).
9. Yoshida, Y. et al. Visualization of patterns of lymph node metastases in non-small cell lung cancer using network analysis. *JTCVS Open* **12**, 410–425 (2022).
10. Agrawal, R., Imielinski, T. & Swami, A. Mining association rules between sets of items in large databases. in *ACM SIGMOD Conference* 207–216 (1993).
11. Abnet, C. C., Arnold, M. & Wei, W. Q. Epidemiology of esophageal squamous cell carcinoma. *Gastroenterology* **154**(2), 360–373 (2018).
12. Shim, Y. M., Kim, H. K. & Kim, K. Comparison of survival and recurrence pattern between two-field and three-field lymph node dissections for upper thoracic esophageal squamous cell carcinoma. *J. Thorac. Oncol.* **5**(5), 707–712 (2010).
13. Mine, S. et al., *Japanese Classification of Esophageal Cancer, 12th Edition: Part I.* (Esophagus, 2024).

14. Lee, J. & Kim, S. Exploring the role of social networks in affective organizational commitment: Network Centrality, strength of ties, and structural holes. *Am. Rev. Public Admin.* **41**(2), 205–223 (2010).
15. Opsahl, T., Agneessens, F. & Skvoretz, J. Node centrality in weighted networks: Generalizing degree and shortest paths. *Soc. Netw.* **32**(3), 245–251 (2010).
16. Stilou, S. et al. Mining association rules from clinical databases: An intelligent diagnostic process in healthcare. *Stud. Health Technol. Inform.* **84**(Pt 2), 1399–1403 (2001).
17. Matsuda, S. et al. Lymph node metastatic patterns and the development of multidisciplinary treatment for esophageal cancer. *Dis. Esophagus* <https://doi.org/10.1093/dote/doad006> (2023).
18. Li, B. et al. Prevalence of lymph node metastases in superficial esophageal squamous cell carcinoma. *J. Thorac. Cardiovasc. Surg.* **146**(5), 1198–1203 (2013).
19. Kang, C. H. et al. Lymphadenectomy extent is closely related to long-term survival in esophageal cancer. *Eur. J. Cardiothorac. Surg.* **31**(2), 154–160 (2007).
20. Ma, L. et al. Characteristics and clinical significance of recurrent laryngeal nerve lymph node metastasis in esophageal squamous cell carcinoma. *J. BUON* **22**(6), 1533–1539 (2017).
21. Yan, H. J. et al. Preoperative clinical characteristics predict recurrent laryngeal nerve lymph node metastasis and overall survival in esophageal squamous cell carcinoma: A retrospective study with external validation. *Front. Oncol.* **12**, 859952 (2022).
22. Chen, C. et al. Risk factors for lymph node metastasis of the left recurrent laryngeal nerve in patients with esophageal squamous cell carcinoma. *Ann. Transl. Med.* **9**(6), 476 (2021).
23. Matsuda, S. et al. Three-field lymph node dissection in esophageal cancer surgery. *J. Thorac. Dis.* **9**(Suppl 8), S731–S740 (2017).
24. Matsuda, S. et al. Three-field lymph node dissection in esophageal cancer surgery. *J. Thoracic. Dis.* **9**(Suppl 8), S731 (2017).
25. Mizutani, M. et al. Anatomy of right recurrent nerve node: Why does early metastasis of esophageal cancer occur in it?. *Surg. Radiol. Anat.* **28**, 333–338 (2006).
26. Li, H. et al. Thoracic recurrent laryngeal lymph node metastases predict cervical node metastases and benefit from three-field dissection in selected patients with thoracic esophageal squamous cell carcinoma. *J. Surg. Oncol.* **105**(6), 548–552 (2012).
27. Miyata, H. et al. A prospective trial for avoiding cervical lymph node dissection for thoracic esophageal cancers, based on intraoperative genetic diagnosis of micrometastasis in recurrent laryngeal nerve chain nodes. *J. Surg. Oncol.* **93**(6), 477–484 (2006).

Acknowledgements

This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (Ministry of Science and ICT; No.2022R1A2C209310611) and a grant of the Korean Cancer Survivors Healthcare R&D Project through the National Cancer Center, funded by the Ministry of Health & Welfare, Republic of Korea (RS-2023-CC139856). The funders played no role in study design, data collection, analysis and interpretation of data, or the writing of this manuscript. The authors extend their appreciation to the physicians at SMC's Department of Thoracic and Cardiovascular Surgery for their assistance in collecting patient data for analysis.

Author contributions

SB managed data preprocessing, network analysis, association rule mining, result interpretation, and manuscript drafting and revision. KK designed and supervised the study, interpreted the results, and revised the manuscript. SYP conceived the idea, supervised the study, drafted and revised the manuscript. YJJ, JL, JHC, HKK, YSC, JIZ, and YMS provided patient data and revised the manuscript. All authors have read and approved the final manuscript for submission.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-89340-2>.

Correspondence and requests for materials should be addressed to K.K. or S.Y.P.

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