

Original Article

Application of Network Analysis and Association Rule Mining for visualizing the Lymph Node Metastasis Patterns in Esophageal Squamous Cell Carcinoma

Sangwon Baek, BA¹, Kyunga Kim, PhD^{2,3,4†}, Seong Yong Park, MD, PhD.^{5†}, Yeong Jeong Jeon, MD, PhD⁵, Junghee Lee, MD, PhD⁵, Jong Ho Cho, MD, PhD⁵, Hong Kwan Kim, MD, PhD⁵, Yong Soo Choi, MD, PhD⁵, Jae Il Zo, MD, PhD⁵, Young Mog Shim, MD, PhD⁵

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

[†]Drs. Kyunga Kim and Seong Yong Park contributed equally as cocorresponding authors.

Abstract

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. 1,181 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/L 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